Powerful and Serial Correlation Robust Tests of the Economic Convergence Hypothesis*

Özgen Saygınsoy†
University at Albany - SUNY

January 3, 2004

Abstract

In this paper, a likelihood ratio approach is taken to derive a test of the economic convergence hypothesis in the context of the linear deterministic trend model. The test is designed to directly address the nonstandard nature of the hypothesis, and is a systematic improvement over existing methods for testing convergence in the same context. The test is first derived under the assumption of Gaussian errors with known serial correlation. However, the normality assumption is then relaxed, and the results are naturally extended to the case of covariance stationary errors with unknown serial correlation. The test statistic is a continuous function of individual t-statistics on the intercept and slope parameters of the linear deterministic trend model, and therefore, standard heteroskedasticity and autocorrelation consistent estimators of the long-run variance can be directly implemented. Building upon the likelihood ratio framework, concrete and specific tests are recommended to be used in practice. The recommended tests do not require the knowledge of the form of serial correlation in the data, and they are robust to highly persistent serial correlation, including the case of a unit root in the errors. The recommended tests utilize the nonparametric kernel variance estimators, which are analyzed using the fixed bandwidth (fixed-b) asymptotic framework recently proposed by Kiefer and Vogelsang (2003). The fixed-b framework makes possible the choice of kernel and bandwidth that deliver tests with maximal asymptotic power within a specific class of tests. It is shown that when the Daniell kernel variance estimator is implemented with specific bandwidth choices, the recommended tests have asymptotic power close that of the known variance case, as well as good finite sample size and power properties. Finally, the newly developed tests are used to investigate economic convergence among eight regions of the United States (as defined by the Bureau of Economic Analysis) in the post-World-War-II period. Empirical evidence is found for convergence in three of the eight regions.

Keywords: Likelihood Ratio, Economic Convergence, β-convergence Hypothesis, Joint Inequality, HAC Estimator, Fixed-b Asymptotics, Power Envelope, Unit Root, Linear Trend, BEA Regions.

*I am indebted to Tim Vogelsang for his invaluable comments and suggestions. I also thank Yi Wen, Yongmiao Hong, Nick Kiefer, Nigar Hashimzade and seminar participants at Cornell University, University at Albany - SUNY, Louisiana State University and Oregon State University for helpful comments.

†Please send correspondence to: Özgen Saygınsoy, Department of Economics, BA123-D, University at Albany -SUNY, 1400 Washington Avenue, Albany, NY 12222; Phone: (518) 442-3175; Fax: (518) 442-4733; E-mail: osayginsoy@albany.edu; Web: http://www.albany.edu/~os145785

1. Introduction

The neoclassical growth model, as presented by Solow (1956), predicts that differences in per-capita real incomes among economies with similar steady-state parameters, such as savings rates and human capital growth rates, must be transitory. From a cross-section perspective, this neoclassical notion of economic convergence implies that, after controlling for differences in steady-state characteristics, the correlation between the initial levels of real per-capita incomes and growth rates must be negative. This is referred to as β-convergence, or conditional convergence [see, for example, Barro (1991), and Mankiw, Romer, and Weil (1992)].

In time-series applications, a dynamic version of this definition has been adopted. Bernard and Durlauf (1995, 1996) define economic convergence as the time-series of differences in real incomes per capita being zero-mean stationary. Bernard and Durlauf's version of the definition is stricter than the cross-section version which only requires that the income differences be narrowing over time. In an attempt to bridge the gap between the two versions of the definition, Hobijn and Franses (2000) require that the time-series of differences in real incomes per capita be level stationary for economic convergence to have taken place. On the other hand, working with United States (US) regional time-series data, Carlino and Mills (1993) , and Tomljanovic and Vogelsang (2002) model the log of per-capita income in one region relative to that of the national average with a linear deterministic trend model. A region is then said to satisfy the β-convergence hypothesis if opposite signs on the intercept and slope parameters of the trend model can be statistically verified; in other words, a region with an initial per-capita income lower than the national average must exhibit a growth trend which is positively steeper than that of the national average; or vice versa. The authors base their decisions regarding β-convergence of a region on two individual one-sided t-tests on the intercept and slope parameters of the linear trend model.

As defined in the context of a linear trend model, β-convergence hypothesis poses a nonstandard testing problem. The composite inequality statements that define β-convergence naturally map into a union of two disjoint spaces in $\mathbb{R}^2$ and a testing framework of this nature has not been analyzed in the context of a regression model. Yancey, Judge, and Bock (1981) discuss methods of testing whether a subset of the parameter vector in the linear regression model lies in the positive orthant. Their results depend on the restrictive assumptions that columns of the regressor matrix are orthonormal and that the innovations are independent and identically distributed (i.i.d.). Gouriéroux *et al.*(1982) also use the linear regression model and examine likelihood ratio (LR), Wald, and Kuhn-Tucker multiplier tests of inequality constraints on the parameters of the model. They show that the common asymptotic distribution of the three tests is a weighted sum of independent chi-square distributions when the errors are mean-zero Gaussian with a known variance-covariance matrix. Similarly, Wolak (1989) also examines tests of inequality constraints which he generalizes to the linear simultaneous equations model. None of the aforementioned studies, however, use a framework that is suitable to analyzing the composite inequality testing problem posed by the β-convergence hypothesis.

In this paper, a LR approach is taken to derive a test of the β-convergence hypothesis in the context of the linear deterministic trend model. The test is designed to directly address the nonstandard nature of the hypothesis, and is a systematic improvement over existing methods for testing convergence in the same context. The test is first derived under the assumption of Gaussian errors with known serial correlation. However, the normality assumption is then relaxed and the results are naturally extended to the case of covariance stationary errors with unknown serial correlation. The test statistic is a continuous function of individual t-statistics on the intercept and slope parameters of the linear trend model, and therefore, standard heteroskedasticity and autocorrelation consistent (HAC) estimators of the long-run variance can be directly implemented. Building upon the LR framework, concrete and specific tests are recommended to be used in practice. The recommended tests do not require the knowledge of the form of serial correlation in the data, and they are robust to highly persistent serial correlation, including the case of a unit root in the errors. The recommended tests utilize the nonparametric kernel variance estimators, which are analyzed using the fixed-b asymptotic framework recently proposed by Kiefer and Vogelsang (2003). The fixed-b framework makes possible the choice of kernel and bandwidth that deliver tests with maximal asymptotic power within a specific class of tests. It is shown that when the Daniell kernel variance estimator is implemented with specific bandwidth choices, the recommended tests have asymptotic power close that of the known variance case, as well as good finite sample size and power properties.

The remainder of the paper is organized as follows: Section 2 provides a short review of the economic convergence literature from both cross-section and time-series perspectives. The limitations of cross-section methods, as well as the conservative findings of time-series methods are discussed. In Section 3, the β-convergence hypothesis is formally defined as nonstandard restrictions on the intercept and slope parameters of the linear deterministic trend model. A LR framework is developed to specifically address the nonstandard nature of the hypothesis, and a LR test is systematically derived. The test is then extended to the case where the errors are covariance stationary with unknown serial correlation, and the standard HAC estimators of the long-run error variance are implemented. In order to make the test statistic robust to highly persistent serial correlation and a unit root in the errors, a more comprehensive version of Vogelsang's (1998) scaling procedure is described and implemented. Section 4 establishes the limiting distributions of test statistics using the fixed-b asymptotic framework, and describes methods used in computing asymptotic critical values. In Section 5, the fixed-b asymptotic distributions of tests under local alternatives are established, and comparisons of local asymptotic power are made for a wide range of kernels and bandwidths. Concrete and specific recommendations are made for the kernel and bandwidth to be used in practice, based on the power performance of tests. Section 6 uses Monte

Carlo simulation methods to explore the finite sample properties of the recommended tests. In Section 7, the newly developed tests are used to investigate economic convergence among eight regions of the United States (as defined by the Bureau of Economic Analysis (BEA)) in the post-World-War-II (WWII) period. Concluding comments are given in Section 8. Supplemental results, proofs, tables and figures are collected in an appendix.

2. Empirical Tests of Economic Convergence

This section summarizes cross-section and time-series methods of testing economic convergence. The papers cited here certainly do not constitute an exhaustive list of the literature and are intended as examples to which the interested reader could refer and expand upon.

Following the neoclassical growth model prediction that differences in per-capita output among economies with similar steady-state parameters must be transitory, one line of empirical research has emerged that concentrates on testing whether countries worldwide have been converging in terms of real incomes per capita within their convergence "clubs," where a "club" has been loosely defined by a group of countries with similar steady-state parameters, such as savings rates, population growth rates, etc. The basic idea is that within a convergence club poorer countries are expected to grow faster than richer countries as ground-breaking technology (i.e., capital) and wisdom (i.e., labor) tend to get transferred across borders at a relatively fast pace. Along this line of reasoning, the question of convergence can be examined for regions within the US, where a region can be defined to include a number of states with certain similar demographic and/or geographic characteristics. With relatively unrestricted labor and capital transfer possibilities across the states' borders, it is no surprise that regions of the US are generally viewed to fit well the definition of a club.

There have been two main approaches to empirically analyzing economic convergence. The first approach employs a cross-section analysis of a set of economic regions where usually a cross-section regression of annual average growth rates on initial levels of real income per capita and other growth-related control variables is estimated. A negative coefficient on the initial levels variable is then taken as an indication of β-convergence[1]. Cross-section tests constructed in this manner generally reject the null hypothesis of no convergence for clubs of highly industrialized countries [see, for example, Baumol (1986)], as well as US regions [see, for example, Barro and Sala-i-Martin (1992)]. After controlling for other growth-related variables, cross-sectional convergence has been found for some large groups of countries as well [see, for example, Barro (1991), and Mankiw, Romer, and Weil (1992)].

Beginning in the first half of 1990s, researchers started noticing the weaknesses of cross-section

[1] Another notion related to β-convergence is σ-convergence, which is said to occur if cross-sectional dispersion of per-capita incomes declines over time. Friedman (1992) , and Quah (1993) observe that β-convergence does not imply σ-convergence.

methods when used for testing convergence in groups. Reliable inference necessitates existence of data on a large number of economic regions, consequently, study of relatively small convergence clubs is impaired by the limited sample size. Further, cross-section tests cannot differentiate between converging and diverging economies in a given group [see, Quah (1996) for a discussion]. These limitations have encouraged researchers to look for alternative methods, and time series techniques for testing convergence have naturally developed.

One time-series-based definition of convergence, known in the literature as asymptotically perfect convergence (APC), is due to Bernard and Durlauf (1995). APC holds if time series of differences in real incomes per capita between two economies contains neither a unit root nor a time trend so that, asymptotically, these differences converge to zero (*zero-mean stationarity*). This definition of convergence, however, is rather strict. Independent of the current and past levels of per-capita incomes, APC necessitates that the *long-run forecast* of the expectation of per-capita output differences be equal to zero. Consequently, the notion of APC is not useful if the question of interest is whether convergence has been occurring in the *past* observed data. Baumol *et al.* (1994) support this observation by arguing that the degree of convergence need not be perfect in the sense implied by APC, but that convergence might stop once economies under consideration have come relatively close to each other. Barro and Sala-i-Martin (1995) present a theoretical model supporting this view by showing that imitating countries can never completely catch-up to the innovating countries due to the costs of imitation.[2] Hobijn and Franses (2000) further explore the Baumol *et al.* (1994) argument, and they introduce the notion of asymptotically relative convergence (ARC) by noting that time series of differences in real incomes per capita need not converge to zero but to a finite constant (*level stationarity*).

An alternative to APC and ARC is the idea that, while two economies may not have yet converged, they could be on a path towards convergence. This idea was first empirically tested by Carlino and Mills (1993) . Working with US regions, Carlino and Mills (1993) constructed annual time-series of the log of the ratio of regional per-capita incomes to the national average income. They fit a simple linear trend model to the time-series of log-ratios, and argued that two conditions need to hold for convergence taking place: (*i*) shocks to relative per-capita incomes must be temporary (*i.e.*, a unit root in the errors must be rejected–*stochastic convergence*); and, (*ii*) regions with initial per-capita incomes lower than the initial national average per-capita income must be catching up to the national average over time, or vice versa (*i.e.*, the slope and intercept parameters of the linear deterministic trend model must have opposite signs–*dynamic β-convergence*).

The conclusions of time series studies have mostly been in contradiction with those of the cross-section studies. Bernard and Durlauf (1991, 1995), for example, failed to find stationarity in per-capita income differences within various sets of world economies, and hence, no support for

[2]This notion of convergence is sometimes referred to as *convergence as catching-up*.

convergence. Similarly, Brown, Coulson, and Engle (1990) report no time-series evidence backing stochastic convergence among a number of US states. However, in contrast to the results of Bernard and Durlauf (1995) , Linden (2000) finds convergence for the majority of countries over the period 1900-87 by using nonparametric tests based on signs and ranks of time-series properties of output differences. Towards explaining inconsistencies in findings between the two methods, Bernard and Durlauf (1996) emphasize that β-convergence and stationarity-type tests must produce conflicting results when applied to the same data sets because the two kinds of approaches explore fundamentally different perspectives of the data at hand. They argue that cross-section methods are well-suited for analyzing economies in transition where initial conditions are important, while stationarity-based time-series methods require economies under analysis to be near their long-run equilibria so that transitory dynamics that might invalidate stationarity tests are not persistent.

Given the apparent difficulties in identifying and controlling for transitory information in data sets, the time-series research have concentrated on testing convergence among US regions where free trade and highly mobile factors among states imply that most conditions underlying convergence must be relatively better satisfied than any other set of economies in the world. If cross-sectional β-convergence results for the US regions cannot be backed by time-series methods, then what hope would there be for establishing such consistency for other world countries?

Carlino and Mills (1993) found evidence of β-convergence in six of the eight US regions for the pre-WWII period (1929-1946), and in four of the eight regions for the post-WWII period (1947-1990). Loewy and Papell (1996) strengthened the results of Carlino and Mills (1993) with respect to stochastic convergence and showed that if break dates are determined endogenously, stochastic convergence is found in seven of the regions. On the other hand, Tomljanovic and Vogelsang (2002) checked the robustness of Carlino and Mills' (1993) results with respect to β-convergence and extended them to the case of an unknown break date by using unit-root robust econometric tests. They reported evidence in favor of β-convergence in six regions in the period following the endogenously determined break date.

One problem associated with the β-convergence testing procedure employed by Carlino and Mills (1993) and Tomljanovic and Vogelsang (2002) is that they use results from a sequence of tests to make conclusions regarding one composite inequality hypothesis, which leaves the overall size of the test unknown. In the next section, a LR framework is developed to specifically address this nonstandard nature of the β-convergence hypothesis. The developed framework will then be used to investigate β-convergence among US regions in the post-WWII period (1947-2002).

3. The Econometric Model

3.1. The LR Test

Suppose there is a set of economic regions among which the β-convergence hypothesis is to be tested. For a given region, let $\{y_t\}_{t=1}^{T}$ denote the natural logarithm of the ratio of per-capita real income to cross-regional average per-capita real income, over T time periods.[3] At time t, y_t is modeled as having a linear deterministic trend function,

$$y_t = \beta_1 + \beta_2 t + u_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where β_1 is the initial level of $\{y_t\}$, *i.e.* the natural logarithm of the initial level of income-to-average-income ratio, and β_2 is the average growth of $\{y_t\}$, *i.e.* the average growth *rate* of income-to-average-income ratio.[4] Let $\mathbf{u} = (u_1, u_2, \ldots, u_T)'$ denote the vector of innovations. Represent the j^{th} autocovariance function of $\{u_t\}$ with $\gamma_j = Cov\left(u_t, u_{t-j}\right)$. For the purpose of deriving a LR test, the following assumption on the distribution of $\{u_t\}$ is made:

Assumption 1: $\mathbf{u} \sim$ *covariance stationary* $\mathbb{N}\left(\mathbf{0}, \mathbf{\Sigma}\right)$, *where* $\mathbf{\Sigma}_{ij} = \gamma_{|i-j|}$ *is the* $(i,j)^{th}$ *entry of the variance-covariance matrix* $\mathbf{\Sigma}$.

Note that, later in this section, it will be possible to naturally extend the test to the case where the distribution of $\mathbf{u}$ is unknown.

The convergence hypothesis, as defined in the context of model (1), necessitates that for regions where y_t is initially positive (*i.e.* initial income is larger than initial average income so that their ratio is larger than 1, and thus, the log of the ratio is positive), the average growth of y_t is negative (*i.e.* income grew, on average, slower than the average of incomes combined so that their ratio is between zero and one, and thus, the log of the ratio is negative); or vice versa. Consequently, for β-convergence hypothesis to be satisfied, it is necessary that if $\beta_1 < 0$ then $\beta_2 > 0$, and, if $\beta_1 > 0$ then $\beta_2 < 0$. But, if convergence has already occurred, then $\beta_1 = \beta_2 = 0$. This is clearly a nonstandard

[3] Throughout the paper, whenever the word "income" is used, it is intended to mean the "real income per capita," unless otherwise stated.

[4] A similar but alternative model design could be implemented if one wishes to test whether two given regions are converging together: Let $\{y_{1t}\}_{t=1}^{T}$ and $\{y_{2t}\}_{t=1}^{T}$ be two time series of observations on natural logs of incomes of the two regions. For $t = 1, 2, \ldots, T$, consider the two models:

$$y_{1t} = \beta_{11} + \beta_{12} t + u_{1t},$$
$$y_{2t} = \beta_{21} + \beta_{22} t + u_{2t},$$

where β_{12} and β_{22} are the average growth rates of incomes, while β_{11} and β_{21} provide measures of the initial incomes. Define $y_t = y_{1t} - y_{2t}$, $\beta_1 = \beta_{11} - \beta_{21}$, $\beta_2 = \beta_{12} - \beta_{22}$, and $u_t = u_{1t} - u_{2t}$, and consider the model,

$$y_t = \beta_1 + \beta_2 t + u_t.$$

Now, techniques described in the paper directly apply to this model for testing β-convergence between the regions.

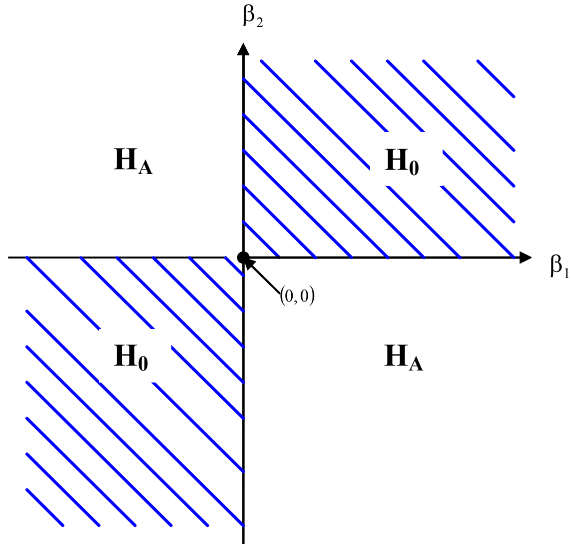


Figure 1: The null and alternative hypotheses.

testing problem that is similar in nature to tests analyzed by Chernoff (1954), Gouriéroux *et al.* (1982), Feder (1968) , Kudô (1963) , Perlman (1969) , Self and Liang (1987) , Wolak (1989) , and others. However, this testing problem does not naturally fit any of these previous frameworks.

It is convenient to adopt the null hypothesis to be that "β-convergence has not been occurring," and the alternative as "β-convergence is occurring." As shown in Figure 1, the null hypothesis can be easily illustrated in the two-dimensional parameter space. The null space is described by the first and third quadrants, including the axes and the origin. The alternative space is the complement of the null space. The null and alternative hypotheses are now formally defined as follows:

$$\underset{[\beta-convergence\ has\ NOT\ been\ occuring]}{H_0 : \boldsymbol{\beta} \in \boldsymbol{\omega}_0} \qquad v. \qquad \underset{[\beta-convergence\ is\ occuring]}{H_A : \boldsymbol{\beta} \in \boldsymbol{\omega}_A} \tag{2}$$

where

$$\boldsymbol{\omega}_0 = \left\{(\beta_1, \beta_2) \in \mathbb{R}^2 : (\beta_1 \leq 0\,,\ \beta_2 \leq 0) \cup (\beta_1 \geq 0\,,\ \beta_2 \geq 0)\right\}, \tag{3}$$

$$\boldsymbol{\omega}_A = \left\{(\beta_1, \beta_2) \in \mathbb{R}^2 : (\beta_1 < 0\,,\ \beta_2 > 0) \cup (\beta_1 > 0\,,\ \beta_2 < 0)\right\}. \tag{4}$$

Having "no convergence" under the null hypothesis is not new to the literature. Indeed, it aligns well with stochastic convergence tests where the null hypothesis of "no convergence" corresponds to a unit root in time series of per-capita income differences. Note that, under the null hypothesis, the case where convergence has already occurred, $\beta_1 = \beta_2 = 0$, is also included. This particular case where convergence has taken place can easily be checked by using an F-type statistic and testing joint equality to zero of the regression parameters in (1).

The likelihood ratio test is first derived under the assumption that $\mathbf{\Sigma}$ is known. The case of unknown $\mathbf{\Sigma}$ is discussed subsequently. Let $L(\mathbf{X}, \boldsymbol{\beta})$ be the likelihood function, where $\mathbf{X}$ is the regressor matrix associated with (1) and $\boldsymbol{\beta}$ is the corresponding 2×1 parameter vector. Under Assumption 1, the likelihood is given by the joint density of the innovations,

$$\begin{aligned} L(\mathbf{X}, \boldsymbol{\beta}) = f(\mathbf{u}) &= (2\pi)^{-T/2} |\mathbf{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u}\right] \\ &= (2\pi)^{-T/2} |\mathbf{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]. \end{aligned} \tag{5}$$

In order to construct a test of the hypotheses specified in $(2)-(4)$, define the likelihood ratio as,

$$\lambda(\mathbf{X}) = \frac{\sup_{\boldsymbol{\beta} \in \boldsymbol{\omega}_0} L(\mathbf{X}, \boldsymbol{\beta})}{\sup_{\boldsymbol{\beta} \in (\boldsymbol{\omega}_0 \cup \boldsymbol{\omega}_A)} L(\mathbf{X}, \boldsymbol{\beta})}. \tag{6}$$

Note that $\lambda \in [0,1]$, and the null hypothesis is rejected for small values of λ. Using (5) and (6), it is straightforward to show that the LR test statistic for the hypotheses in $(2)-(4)$ is proportional to

$$LR \equiv -2\log\lambda(\mathbf{X}) = \inf_{\boldsymbol{\beta} \in \boldsymbol{\omega}_0} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) - \inf_{\boldsymbol{\beta} \in (\boldsymbol{\omega}_0 \cup \boldsymbol{\omega}_A)} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y}-\mathbf{X}\boldsymbol{\beta}). \tag{7}$$

Define the unconstrained and constrained maximum likelihood (ML) estimators for the regression parameter vector $\boldsymbol{\beta}$ in (1) respectively as,

$$\widehat{\boldsymbol{\beta}}_{ml} \equiv \arg \sup_{\boldsymbol{\beta} \in (\boldsymbol{\omega}_0 \cup \boldsymbol{\omega}_A)} L(\mathbf{X}, \boldsymbol{\beta}) = \left(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{y}, \tag{8}$$

and,

$$\widehat{\boldsymbol{\beta}}_0 \equiv \arg \sup_{\boldsymbol{\beta} \in \boldsymbol{\omega}_0} L(\mathbf{X}, \boldsymbol{\beta}). \tag{9}$$

Using (8) and (9), the expression for LR given in (7) can be written as,

$$LR = \left[\left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_0\right)' \mathbf{\Sigma}^{-1} \left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_0\right) - \left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{ml}\right)' \mathbf{\Sigma}^{-1} \left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{ml}\right)\right]. \tag{10}$$

By utilizing the expression $\mathbf{y}-\mathbf{X}\boldsymbol{\beta} = \left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{ml}\right) + \mathbf{X}\left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right)$ and the first order conditions from the unconstrained ML estimation, it is easily shown that

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{ml}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}}_{ml}\right) + \left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right)'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right). \tag{11}$$

Finally, by using (11), the expression in (10) can be rewritten as follows:

$$LR = \left(\widehat{\boldsymbol{\beta}}_{ml}-\widehat{\boldsymbol{\beta}}_0\right)'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\widehat{\boldsymbol{\beta}}_{ml}-\widehat{\boldsymbol{\beta}}_0\right) = \inf_{\boldsymbol{\beta}\in\boldsymbol{\omega}_0}\left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right)'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right). \tag{12}$$

This result transforms the problem from T dimensions down to 2 dimensions, and is summarized in the following theorem.

Theorem 1 *If Assumption 1 holds, then the LR test statistic for testing the null hypothesis specified in* (2) – (4) *for* $\boldsymbol{\beta}$ *in model* (1) *is equivalent to the LR test statistic for testing the same hypotheses for* $\boldsymbol{\beta}$ *in model* $\widehat{\boldsymbol{\beta}}_{ml} = \boldsymbol{\beta} + \boldsymbol{\upsilon}$, *where* $\widehat{\boldsymbol{\beta}}_{ml}$ *is the unconstrained ML estimator of* $\boldsymbol{\beta}$ *and* $\boldsymbol{\upsilon}$ *is distributed as* $\mathbb{N}\left[\mathbf{0}, \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\right]$.

The result given in Theorem 1 is not completely new. Under the assumption of i.i.d. errors, Chernoff (1954) showed that a result analogous to Theorem 1 holds if the null and alternative spaces are disjoint subsets of the Euclidean k-space, the true parameter vector is a boundary point of both the null and alternative spaces, and both spaces are approximable at the true parameter vector (without loss of generality, taken as equal to $\mathbf{0}$) by positively homogeneous sets (cones). Feder (1968), and Self and Liang (1987) generalize Chernoff's results to cases where the true parameter vector is not a boundary point of the null and alternative spaces. On the other hand, Gouriéroux *et al.* (1982) derive and use a result similar to Theorem 1, but to test composite inequality hypotheses of a different nature.

Now that the LR statistic is configured in a more manageable way by (12), its exact form can be explicitly solved. Let $\widehat{\boldsymbol{\beta}}_{ml} = \left[\widehat{\beta}_{1,ml} \;\; \widehat{\beta}_{2,ml}\right]'$, and define,

$$\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} \equiv \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{22}^2 \end{bmatrix}, \tag{13}$$

and,

$$\rho_{12} \equiv Corr\left(\widehat{\beta}_{1,ml}\,,\,\widehat{\beta}_{2,ml}\right) = \frac{\sigma_{12}}{\sqrt{\sigma_{11}^2\sigma_{22}^2}} = \frac{\sigma_{12}}{\sigma_{11}\sigma_{22}}. \tag{14}$$

Using (13) and (14), it follows by straightforward algebra that $\left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right)'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\widehat{\boldsymbol{\beta}}_{ml}-\boldsymbol{\beta}\right)$ is equal to,

$$\frac{1}{1-\rho_{12}^2}\left[\left(\frac{\widehat{\beta}_{1,ml}-\beta_1}{\sigma_{11}}\right)^2+\left(\frac{\widehat{\beta}_{2,ml}-\beta_2}{\sigma_{22}}\right)^2-2\rho_{12}\left(\frac{\widehat{\beta}_{1,ml}-\beta_1}{\sigma_{11}}\right)\left(\frac{\widehat{\beta}_{2,ml}-\beta_2}{\sigma_{22}}\right)\right]. \tag{15}$$

For the intercept and slope parameters of model (1) respectively, define the individual t-tests, $t_i = \left(\widehat{\beta}_{i,ml}-\beta_i\right)\Big/\sigma_{ii}$ for $i=1,2$. Using (15), (12) can be re-expressed as,

$$LR = \frac{1}{1-\rho_{12}^2}\inf_{(t_1,t_2)\in\omega_0'}\left[t_1^2+t_2^2-2\rho_{12}t_1t_2\right], \tag{16}$$

where,

$$\omega_0' = \left\{(t_1,t_2)\in\mathbb{R}^2 : \left(t_1 \geq \frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\ ,\ t_2 \geq \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\right)\cup\left(t_1 \leq \frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\ ,\ t_2 \leq \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\right)\right\} \tag{17}$$

The expression in (16) is an elliptic paraboloid or ellipsoid in (t_1,t_2)-space, and hence, the solution is directly attainable by methods of multivariate calculus. The following theorem provides the form of the LR test in (16). The proof can be found in the appendix.

Theorem 2 *Suppose that Assumption 1 holds. Then, under the null hypothesis specified in* $(2)-(4)$*, the LR test statistic is,*

$$LR = 1\left(t_1t_2<0\right)\cdot\min\left[t_1^2\ ,\ t_2^2\right], \tag{18}$$

where $1(\cdot)$ *is the indicator function, and* $t_i = \widehat{\beta}_{i,ml}\Big/\sigma_{ii}$ *for* $i=1,2$.

The LR test statistic presented in Theorem 2 does not reject the null hypothesis when t_1 and t_2 have the same sign. If t_1 and t_2 have opposite signs, the test evaluates the magnitude of deviations from the part of the null in favor of which there is greater evidence. Suppose, for example, that $t_1<0$ and $t_2>0$. If $|t_1|>t_2$, then there is more evidence in favor of the part of the null given by $\{(\beta_1,\beta_2)\in\mathbb{R}^2:\beta_1\leq 0\ ,\ \beta_2\leq 0\}$ than the other part, given by $\{(\beta_1,\beta_2)\in\mathbb{R}^2:\beta_1\geq 0\ ,\ \beta_2\geq 0\}$. The test checks whether the t-statistic not satisfying the favored part of the null hypothesis, in our example t_2, is large enough so that the null hypothesis can be rejected.

3.2. The *Pseudo*-LR Test

The result in Theorem 2 is based on the ML estimates of the regression parameters and the assumption of a known variance-covariance matrix of the errors. This result, however, can be extended to the case where the errors are covariance stationary but with an unknown variance-covariance matrix.

It is implicit in Theorem 1 that, under Assumption 1, the ML estimates of regression parameters in model (1) are asymptotically equivalent to the GLS estimates. The classic results of Grenander and Rosenblatt (1957) ensure that under covariance stationary errors, OLS estimates of regression parameters remain efficient, because they are asymptotically normal, with variance equivalent to that of GLS. Consequently, the ML estimates in the definition of the LR statistic introduced in Theorem 2 can be replaced with the OLS estimates, and the test statistic would still remain asymptotically valid.

To state this fact more formally, let $\widehat{\boldsymbol{\beta}} = \left[\widehat{\beta}_1 \;\; \widehat{\beta}_2\right]'$ denote the vector of OLS estimates of β_1 and β_2 in model (1). The standard variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ is given by $\sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$. Let $\widehat{\sigma}^2$ be a consistent estimator of σ^2. Then, the *pseudo*-LR (PLR) statistic is defined as,

$$PLR = 1\left(t_1 t_2 < 0\right) \cdot \min\left[t_1^2,\, t_2^2\right], \tag{19}$$

where,

$$t_1 = \frac{\widehat{\beta}_1 - \beta_1}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)_{11}^{-1}\right]^{1/2}}, \qquad\qquad t_2 = \frac{\widehat{\beta}_2 - \beta_2}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)_{22}^{-1}\right]^{1/2}}, \tag{20}$$

are the standard t-statistics on the intercept and slope parameters of (1). Note that $(\mathbf{A})_{jj}^{-1}$ refers to the j^{th} diagonal entry of the inverse of matrix $\mathbf{A}$. Under covariance stationary errors, the PLR statistic, as defined above with the OLS estimates and a consistent estimator of σ^2, is asymptotically equivalent to the LR statistic in (18), which is defined with the ML estimates using a known variance-covariance matrix.

Since all elements of the regressor matrix are known, the limiting distribution of PLR can be explicitly derived by using the joint asymptotic distribution of t_1 and t_2, which is given by,

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \sim \mathbb{N}\left(\mathbf{0}_{2\times 1}, \begin{bmatrix} 1 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & 1 \end{bmatrix}\right). \tag{21}$$

The asymptotic distribution of PLR is presented in the following theorem.

Theorem 3 *Estimate model* (1) *by OLS, define the PLR statistic as in* (19), *and construct the standard t-statistics on the intercept and slope parameters as in* (20) *with a consistent estimator of σ^2. Then, if the errors are covariance stationary, PLR has a probability density function* (*p.d.f.*) *given by,*

$$f_{PLR}(v) = \frac{1}{6} \cdot \delta(v) + \frac{5}{6} \cdot \omega(v) \cdot f_Z(v), \quad 0 \le v < \infty, \tag{22}$$

where, $f_Z(v) = e^{-v/2}/\sqrt{2\pi v}$, $0 \le v < \infty$, is the p.d.f. for a chi-square random variable with 1 *degree of freedom, $\delta(\cdot)$ is the dirac delta function, and,*

$$\omega(v) = 2\left[\phi\left(\left(-2+\sqrt{3}\right)\sqrt{v}\right) - \phi\left(\left(2+\sqrt{3}\right)\sqrt{v}\right)\right],$$

where $\phi(\cdot)$ is the standard normal cumulative distribution function.

The limiting distribution of PLR is nonstandard, however, critical values can easily be computed. The most commonly used critical values have been tabulated in Table 2.

3.4. The PLR Test *vs.* the *Ad Hoc* Procedure

It is interesting to note that the *ad hoc* approach to testing β-convergence used by Carlino and Mills (1993), and Tomljanovic and Vogelsang (2002) is computationally equivalent to the PLR test described above, however, it uses rejection rules that are asymptotically invalid given the nonstandard nature of the underlying problem. Their *ad hoc* inference procedure relies on two one-sided t-tests on the intercept and slope parameters of model (1) and implicitly ignores the effect on asymptotic theory of the nonzero correlation between t_1 and t_2. One problem associated with using results from a sequence of tests to make conclusions regarding a nonstandard, composite inequality hypothesis is that the overall size of the performed test remains unknown.

This observation is illustrated in Table 1, where empirical null hypothesis rejection probabilities are reported for the PLR statistic and the *ad hoc* procedure using nominal test sizes equal to 1%, 2.5%, 5%, 10%, 15% and 20%, and sample sizes equal to 50, 100, 200, 500 and 1000. The data were generated according to (1) with i.i.d. standard normal errors. OLS was used in estimating the regression parameters, and σ^2 was estimated by using $s^2 = \frac{1}{T-2}\sum_{t=1}^{T}\widehat{u}_t^2$. The rejection probabilities for PLR were computed using asymptotic critical values from Table 2. When computing rejection probabilities using the *ad hoc* procedure, there were two steps involved: At the first step, signs of t_1 and t_2 were checked. If they were found to be of opposite signs, then it was proceeded to step 2, otherwise, the null hypothesis was *not* rejected. At step two, the standard normal critical values were used (at the nominal level indicated) to carry out each one sided t-test. The null hypothesis was rejected whenever both of the one-sided t-tests were found significant at the indicated nominal level. 1,000,000 replications were performed for each sample size.

It is clearly evident from Table 1 that the *ad hoc* procedure suffers from size distortions, which get increasingly worse as the nominal level of the test increases, and fail to disappear as the sample size gets large. On the other hand, when tests are carried out using the PLR statistic, empirical rejection probabilities remain close to the corresponding nominal size.

3.5. The PLR Test with Conventional HAC Estimators of σ^2

In order to further examine the finite sample properties of the PLR statistic under various error specifications, it is necessary that a serial correlation robust estimator of σ^2 be used. In this paper, the PLR statistic will be implemented by using the class of nonparametric estimators of the long-run variance, defined by,

$$\widehat{\sigma}^2 = \widehat{\gamma}_0 + 2\sum_{j=1}^{T-1} k\left(j/M\right)\widehat{\gamma}_j, \tag{23}$$

where M is the truncation lag or the bandwidth parameter, $\widehat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T}\widehat{u}_t\widehat{u}_{t-j}$ are the sample autocovariances, $\{\widehat{u}_t\}$ are the residuals from the OLS estimation of (1), and $k(x)$ is a kernel function that is continuous at $x = 0$ and satisfies $k(x) = k(-x)$, $k(0) = 1$, $|k(x)| \leq 1$ and $\int_0^1 k(x)^2\, dx < \infty$. For the consistency of $\widehat{\sigma}^2$, it is necessary that as $T \to \infty$, $M \to \infty$ and $M/T \to 0$. Kernel functions that are used in this paper are listed in the appendix.

One conventional estimator of σ^2, labelled $\widehat{\sigma}^2_{HAC}$, is obtained by using the Bartlett kernel and Andrews' (1991) data-dependent AR(1) "plug-in" formula for determining the bandwidth parameter (M). Andrews (1991) showed that the bandwidth parameter that minimizes the truncated asymptotic mean squared error of $\widehat{\sigma}^2_{HAC}$ should grow at rate $T^{1/3}$ for the Bartlett kernel, and therefore, $\widehat{\sigma}^2_{HAC}$ is a consistent estimator σ^2 (see the original paper for more details).

Den Haan and Levin (1998), on the other hand, recommend a parametric spectral estimation procedure. A univariate version of their method is adopted here. AR models up to the fourth order are fit to the residuals of model (1) and Schwarz (1978) Bayesian Information Criterion (BIC) is used in identifying the best fitting model. Let k denote the AR lag length that maximizes BIC. Consider the fitted regression, $\widehat{u}_t = \widehat{\rho}_1\widehat{u}_{t-1} + \cdots + \widehat{\rho}_k\widehat{u}_{t-k} + \widehat{e}_t$, where $\{\widehat{e}_t\}$ are the OLS residuals. The PARM estimator is then defined as,

$$\widehat{\sigma}^2_{PARM} = \frac{(T-k)^{-1}\sum_{t=k+1}^{T}\widehat{e}_t^2}{\left(1 - \sum_{n=1}^{k}\widehat{\rho}_n\right)^2}. \tag{24}$$

In order to examine the finite sample size performance of PLR with HAC and $PARM$ estimators of σ^2, the data were generated according to model (1) with ARMA$(1,1)$ errors,

$$u_t = \rho u_{t-1} + e_t + \theta e_{t-1}, \tag{25}$$

where $\{e_t\}$ is i.i.d. $\mathbb{N}(0,1)$, $u_0 = 0$, and $e_0 = 0$. OLS is used in estimating the intercept and slope parameters that are used in constructing the tests. For each test, the null hypothesis rejection probabilities were calculated for various values of ρ and θ by using 10% asymptotic null critical values from Table 2. For the HAC estimator, the results are reported without prewhitening and also with AR(1) prewhitening ($HACPW$). The results are reported in Tables 3-5 for sample sizes equal to 50, 100, and 200 respectively. 10,000 replications were performed in each case.

It is evident in Tables 3-5 that regardless of which estimator is used, PLR test suffers from severe size distortions in finite samples even under the slightest presence of AR(1) serial correlation, and regardless of the MA component. When $\widehat{\sigma}^2_{HAC}$ is used, PLR exhibits generally the poorest performance of the three estimators across all sample sizes. On the other hand, $\widehat{\sigma}^2_{PARM}$ delivers relatively the better performance when there are negative MA components present, while the same is true for $\widehat{\sigma}^2_{HACPW}$ when there are positive MA components present. When there is only AR(1) serial correlation present, $\widehat{\sigma}^2_{PARM}$ and $\widehat{\sigma}^2_{HACPW}$ deliver tests with comparable finite sample size performance.

It is clear in Tables 3-5 that when the errors are characterized by strong serial correlation, the standard asymptotic rejection rules used in the stationary case are inaccurate in finite samples and lead to over-rejection problems. For practical purposes, it will be useful if the theory can be strengthened so that size distortions are reduced under persistent serial correlation. This will be achieved by developing the null asymptotic theory of PLR under the fixed-b asymptotic framework, and implementing a more comprehensive version of the scaling factor approach of Vogelsang (1998). However, first, the assumptions on the underlying error structure need to be more explicitly stated.

3.6. The Scaled PLR Tests

The following assumptions on the error process $\{u_t\}$ are used for the remainder of the paper:

Assumption 1′:

$$
\begin{aligned}
u_t &= \alpha u_{t-1} + \varepsilon_t, \qquad t = 2, 3, \ldots, T, \\
u_1 &= \varepsilon_1, \\
\varepsilon_t &= d(L)\, e_t, \qquad d(L) = \sum_{i=0}^{\infty} d_i L^i, \qquad \sum_{i=0}^{\infty} i\,|d_i| < \infty, \qquad d(1)^2 > 0,
\end{aligned}
$$

where $\{e_t\}$ *is a martingale difference sequence with* $E\left(e_t | e_{t-1}, e_{t-2}, \ldots\right) = 0$, $E\left(e_t^2 \middle| e_{t-1}, e_{t-2}, \ldots\right) = 1$ *and* $\sup_t E\left(e_t^4\right) < \infty$. L *is the lag operator.*

The error process $\{u_t\}$ is $I(0)$ when $|\alpha| < 1$. Alternatively, $\{u_t\}$ can be modeled as a nearly $I(1)$ process by defining $\alpha = (1 - \overline{\alpha}/T)$, where $\overline{\alpha} = 0$ corresponds to the pure $I(1)$ case.[5] The restriction $d(1)^2 > 0$ effectively bounds the spectral density of $\{\varepsilon_t\}$ at frequency zero above zero and serves to eliminate nondegenerate cases. Assumption 1′ ensures that the following well-known functional central limit theorem results hold [see, for example, Phillips (1987), and, Phillips and Solo (1992)]:

$$
T^{-1/2} \sum_{t=1}^{[rT]} u_t \implies \sigma W(r), \qquad \text{if } \{u_t\} \text{ is I}(0), \tag{26}
$$

$$
T^{-1/2} u_{[rT]} \implies d(1)\, V_{\overline{\alpha}}(r), \qquad \text{if } \{u_t\} \text{ is I}(1), \tag{27}
$$

[5] Throughout the paper, whenever errors are said to be $I(1)$, it is intended to mean that errors exhibit a unit root locality, unless otherwise stated.

where $[rT]$ denotes the integer part of rT, $r \in [0,1]$, "$\Longrightarrow$" denotes weak convergence, $W(r)$ is the standard Wiener process, $\sigma^2 = d(1)^2 \big/ (1-\alpha)^2$, and $V_{\overline{\alpha}}(r) = \int_0^r \exp(-\overline{\alpha}(r-s))\, dW(s)$.

The basic idea underlying the scaling factor approach is to multiplicatively use an exponential function of a unit root statistic to smooth discontinuities in the asymptotic distributions of test statistics as the errors go from $I(0)$ to $I(1)$. Two unit root statistics will be employed to construct the scaling factors to be used with PLR statistic.

The first unit root statistic, denoted by J, was proposed by Park and Choi (1988) and Park (1990). The J statistic is the standard OLS Wald statistic normalized by T for testing the joint hypothesis $\pi_2 = \pi_3 = \ldots = \pi_9 = 0$ in the regression model,

$$y_t = \beta_1 + \beta_2 t + \sum_{i=2}^{9} \pi_i t^i + u_t. \tag{28}$$

In particular, the J statistic is defined as,

$$J = \frac{RSS_R - RSS_U}{RSS_U},$$

where RSS_U is the residual sum of squares from the "unrestricted" regression in (28),and RSS_R is the residual sum of squares from the "restricted" regression in (1).

The second unit root statistic, denoted by BG, is the variance-ratio statistic of Breitung (2002). The variance-ratio statistic is an LM-type statistic similar to the statistics proposed by Tanaka (1990) and Kwiatkowski *et al.* (1992), and is defined as,

$$BG = \frac{T^{-2} \sum_{t=1}^{T} \widehat{S}_t^2}{RSS_R},$$

where $\widehat{S}_t = \sum_{j=1}^{t} \widehat{u}_t$ are the partial sums of the OLS residuals from the estimation of (1).

Let UR generically denote either J or BG statistic. Let c_{UR} be a constant. The scaling factor is then defined as $\exp(-c_{UR} \cdot UR)$. Using this scaling factor, the scaled PLR statistic is then generically defined as,

$$\begin{aligned} PLR_{UR} &= 1(t_1 t_2 < 0) \cdot \min\left[t_1^2, t_2^2\right] \cdot \exp(-c_{UR} \cdot UR) \\ &= PLR \cdot \exp(-c_{UR} \cdot UR). \end{aligned} \tag{29}$$

Note that both J and BG are left-tailed unit root tests, and thus, they reject the null hypothesis of a unit root in the errors for small values. When the errors are stationary, both J and BG converge to zero, and thus, $\exp(-c_{UR} \cdot UR)$ converges to one and has no effect on the limiting distribution of PLR_{UR}. However, when the errors have a unit root, J and BG have nondegenerate limiting distributions (that are free of nuisance parameters), and therefore, given a significance level, and the local unit root parameter $\overline{\alpha}$, the constant c_{UR} can be chosen so that stationary critical values and unit root critical values (given $\overline{\alpha}$) are the same. In other words, given a significance level, and a value for $\overline{\alpha}$, careful choice of c_{UR} renders PLR_{UR} statistic asymptotically size correct whether errors are stationary or have a near unit root *with* locality parameter $\overline{\alpha}$.

4. Asymptotic Theory and Critical Values

In this section, the asymptotic null distribution theory of PLR_{UR} tests are established under the fixed-b asymptotic framework. The bandwidth of the covariance matrix estimator in (23) is modeled as a fixed proportion of the sample size by letting $M = bT$, where $b \in (0, 1]$. This contrasts the traditional asymptotics where the bandwidth increases slower than the sample size and asymptotic distributions of HAC robust tests do not depend on the bandwidth or the kernel.

Some definitions are required before stating the asymptotic results.

Definition 1. *A kernel is labelled as Type 1 if $k(x)$ is twice continuously differentiable everywhere, and as Type 2 if $k(x)$ is continuous, twice continuously differentiable everywhere except at $|x| = 1$, and $k(x) = 0$ for $|x| \geq 1$.*

Note that Bartlett kernel classifies as neither Type 1 nor a Type 2 kernel, and hence, is considered separately. Definition 2 below intends to simplify notation in the limiting distributions that follow.

Definition 2. *Let $k^*(\cdot) = k(\cdot/b)$, and let $k^{*\prime}_{-}(\cdot)$ denote the first derivative of $k^*(\cdot)$ from below. Denote the second derivative of $k^*(\cdot)$ with $k^{*\prime\prime}(\cdot)$. Define,*

$$\widehat{Q}(r) = \begin{cases} \widehat{Q}_0(r) = W(r) + r(2-3r)W(1) + 6r(r-1)\int_0^1 W(s)\,ds, & \text{if } \{u_t\} \text{ is } I(0) \\ \widehat{Q}_1(r) = V_{\overline{\alpha}}(r) - (4-6r)\int_0^1 V_{\overline{\alpha}}(s)\,ds + (6-12r)\int_0^1 sV_{\overline{\alpha}}(s)\,ds, & \text{if } \{u_t\} \text{ is } I(1) \end{cases}$$

$$\phi(b,k) = \begin{cases} \int_0^1\int_0^1 -k^{*\prime\prime}(r-s)\,\widehat{Q}(r)\,\widehat{Q}(s)\,drds, & \text{if } k(\cdot) \text{ is Type 1} \\ \int\int_{|r-s|<b} -k^{*\prime\prime}(r-s)\,\widehat{Q}(r)\,\widehat{Q}(s)\,drds \; + \\ \qquad 2k^{*\prime}_{-}(b)\int_0^{1-b}\widehat{Q}(r+b)\,\widehat{Q}(r)\,dr, & \text{if } k(\cdot) \text{ is Type 2} \\ \frac{2}{b}\left\{\int_0^1 \widehat{Q}(r)^2\,dr - \int_0^{1-b}\widehat{Q}(r+b)\,\widehat{Q}(r)\,dr\right\}, & \text{if } k(\cdot) \text{ is Bartlett} \end{cases}$$

In the case of $I(1)$ errors, the limiting distributions of test statistics further depend upon the limiting distributions of the unit root statistics, J and BG, that are used to construct the scaling factors. The following lemma establishes some necessary asymptotic results that directly follow from Park (1990), Park and Choi (1988) and Breitung (2002).

Lemma 1 *Let* $V_{\overline{\alpha}}^{*}(r)$ *denote the residuals from the projection of* $V_{\overline{\alpha}}(r)$ *onto the space spanned by* $\left(1,r,r^2,r^3,\ldots,r^9\right)'$ *on* $[0,1]$. *Also, denote the residuals from the projection of* $V_{\overline{\alpha}}(r)$ *onto the space spanned by* $(1,r)'$ *on* $[0,1]$ *by* $\widehat{V}_{\overline{\alpha}}(r)$. *Suppose Assumption* $1'$ *holds. Then, as* $T \rightarrow \infty$, *if* $\{u_t\}$ *is* $I(0)$,

$$J \Rightarrow 0 \quad \text{and} \quad BG \Rightarrow 0,$$

and, if $\{u_t\}$ *is* $I(1)$,

$$\begin{aligned} J &\Rightarrow \frac{\int_0^1 \widehat{V}_{\overline{\alpha}}(r)^2\,dr - \int_0^1 V_{\overline{\alpha}}^{*}(r)^2\,dr}{\int_0^1 V_{\overline{\alpha}}^{*}(r)^2\,dr}, \\ BG &\Rightarrow \frac{\int_0^1 \widehat{Q}_1(r)^2\,dr}{\int_0^1 \widehat{V}_{\overline{\alpha}}(r)^2\,dr}. \end{aligned}$$

In what follows, the limiting distributions above will generically be denoted by $UR_{\overline{\alpha}}^{\infty}$. The fixed-$b$ asymptotic distribution of $\widehat{\sigma}^2$, given in the following lemma, directly follows from Bunzel and Vogelsang (2003).

Lemma 2 *Suppose that Assumption* $1'$ *holds, and let the bandwidth parameter be a fixed proportion of the sample size,* $M = bT$, $b \in (0,1]$. *Then, as* $T \rightarrow \infty$,

$$\begin{aligned} \widehat{\sigma}^2 &\Rightarrow \sigma^2\phi(b,k), & &\text{if } \{u_t\} \text{ is } I(0), \\ T^{-2}\widehat{\sigma}^2 &\Rightarrow d(1)^2\phi(b,k), & &\text{if } \{u_t\} \text{ is } I(1), \end{aligned} \tag{30}$$

where $\phi(b,k)$ *is defined in Definition 2.*

The fixed-b limiting distribution of PLR_{UR} follows by the continuous mapping theorem, and is stated in the following theorem.

Theorem 4 *Suppose that Assumption* $1'$ *holds, and let the bandwidth parameter be a fixed proportion of the sample size,* $M = bT$, $b \in (0,1]$. *Estimate model* (1) *by OLS, construct the standard HAC robust t-statistics on the intercept and slope parameters as in* (20), *and define the* PLR_{UR} *statistic as in* (29). *Then, as* $T \rightarrow \infty$,

$$t_1 \Rightarrow t_{1,\infty} = \begin{cases} -3\left[\frac{1}{3}W(1) - \int_0^1 W(r)\,dr\right] \Big/ \sqrt{\phi(b,k)}, & \text{if } \{u_t\} \text{ is } I(0) \\ \\ 3\left[\frac{1}{3}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right] \Big/ \sqrt{\phi(b,k)}, & \text{if } \{u_t\} \text{ is } I(1) \end{cases} \tag{31}$$

and,

$$t_2 \Rightarrow t_{2,\infty} = \begin{cases} \sqrt{12}\left[\frac{1}{2}W(1) - \int_0^1 W(r)\,dr\right] \Big/ \sqrt{\phi(b,k)}, & \text{if } \{u_t\} \text{ is } I(0) \\ \\ -\sqrt{12}\left[\frac{1}{2}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right] \Big/ \sqrt{\phi(b,k)}, & \text{if } \{u_t\} \text{ is } I(1) \end{cases} \tag{32}$$

and,

$$PLR_{UR} \Rightarrow 1\left(t_{1,\infty}t_{2,\infty} < 0\right) \cdot \min\left[t_{1,\infty}^2\,,\ t_{2,\infty}^2\right] \cdot \exp\left(-c_{UR}UR_{\overline{\alpha}}^{\infty}\right).$$

Theorem 4 illustrates that the PLR_{UR} statistic is asymptotically free of nuisance parameters when the errors are modeled as $I(0)$, and only depends on the unit root locality parameter $\overline{\alpha}$ when the errors are $I(1)$. The dependence of asymptotic distributions on the kernel and bandwidth is through the limiting distribution of $\widehat{\sigma}^2$. Given the kernel, bandwidth, unit root statistic and a percentage point, the constant, c_{UR}, can be computed such that the PLR_{UR} statistic remains at least asymptotically size conservative, if not exactly size correct, across a wide and fine grid of values of $\overline{\alpha}$.

The critical values for test statistics reported in Theorem 4, as well as values of c_{UR}, are straightforward to compute by means of Monte Carlo simulation methods. Critical values for t_1 and t_2 by themselves have been recently tabulated by Bunzel and Vogelsang (2003). In computing critical values for PLR_{UR} tests, i.i.d. standard normal random deviates were used to approximate the Brownian motions in the asymptotic distributions. Integrals were approximated by normalized partial sums of 1000 steps using 10,000 replications.

The analysis in this paper focuses on five popular kernels: Bartlett, Parzen, Bohman, Daniell, and Quadratic Spectral (QS). For each kernel, asymptotic null critical values have been computed for the grid of bandwidths given by $b = 0.02, 0.04, \ldots, 1.0$. Given a significance level (η), a kernel, a bandwidth, and the set of values of $\overline{\alpha}$, given by $\Psi = \{0, 5, 10, \ldots, 50\}$, the constants, c_{UR}, have been computed such that,

$$\sup_{\overline{\alpha}\in\Psi} P\left[\text{reject } H_0 \mid \{u_t\} \text{ is } I(1)\right] = \eta.$$

A finer grid of values for $\overline{\alpha}$ was not considered, because this would be computationally expensive.

Based on the asymptotic power analysis (to be discussed in the next section), the critical values for PLR_{BG} only for the Daniell kernel and bandwidth values equal to 0.02, 0.12 and 0.22 times the sample size have been tabulated in Table 2. The corresponding values for c_{UR} are also provided in parentheses below each critical value.

5. Asymptotic Power Analysis: Optimal Kernel and Bandwidths

This section presents a comprehensive analysis of asymptotic power for the PLR_{UR} tests. Because the tests are size controlled, the fixed-b asymptotic distributions of tests under a local alternative can be used to make power comparisons for a wide range of kernels and bandwidths. Using this analysis, concrete and specific recommendations are made for the kernel and bandwidth to be used in practice.

Given the nonstandard nature of the hypothesis being tested, deviations from the null can be modeled in two different ways. One choice would be to model only the deviations from the null of β_1 as β_2 remains fixed in the alternative space; or vice versa. The other choice would be to model deviations from the null of both β_1 and β_2 simultaneously. In unreported simulations, it was found that more meaningful power comparisons are possible using the latter method, in the sense that the rate at which power approaches one was relatively slower with the former method.

Under the alternative, β_1 is modeled as local to zero with local alternative parameter $d_1 > 0$, and approaches to zero from above at rate $g_1(T)$, while β_2 is also modeled as local to zero but with local alternative parameter $d_2 = -d_1$ and approaches zero at rate $g_2(T)$, where,

$$g_1(T) = \begin{cases} T^{-1/2}, & \text{if } \{u_t\} \text{ is } I(0) \\ T^{1/2}, & \text{if } \{u_t\} \text{ is } I(1) \end{cases} \qquad \text{and} \qquad g_2(T) = \begin{cases} T^{-3/2}, & \text{if } \{u_t\} \text{ is } I(0) \\ T^{-1/2}, & \text{if } \{u_t\} \text{ is } I(1) \end{cases}. \tag{33}$$

In particular, testing is done for $H_0 : \boldsymbol{\beta} \in \boldsymbol{\omega}_0$ versus $H_A : \boldsymbol{\beta} \in \boldsymbol{\omega}_A$, where,

$$\boldsymbol{\omega}_0 = \left\{(\beta_1, \beta_2) \in \mathbb{R}^2 : (\beta_1 \leq 0\,,\, \beta_2 \leq 0) \cup (\beta_1 \geq 0\,,\, \beta_2 \geq 0)\right\}, \tag{34}$$

$$\boldsymbol{\omega}_A = \left\{(\beta_1, \beta_2) \in \mathbb{R}^2 : \beta_1 = d_1 g_1(T)\,,\, \beta_2 = -d_1 g_2(T)\right\}. \tag{35}$$

The limiting distribution of PLR_{UR} under the local alternative follows by the continuous mapping theorem from the limiting distributions of t-tests on the intercept and slope parameters of (1). Note that the long-run variance estimator $\widehat{\sigma}^2$ and unit root statistics J and BG are exactly invariant to the true values of β_1 and β_2. Therefore, the dependence of PLR_{UR} on the local alternative is through the OLS estimates that are used to construct the t-statistics on the intercept and slope parameters. The theorem below presents the limiting distributions of t_1, t_2, and PLR_{UR} under the local alternative.

Theorem 5 *Suppose that Assumption* $1'$ *holds, and let the bandwidth parameter be a fixed proportion of the sample size,* $M = bT$, $b \in (0, 1]$. *Estimate model* (1) *by OLS, construct the standard HAC robust t-statistics on the intercept and slope parameters as in* (20), *and define the* PLR_{UR} *statistic as in* (29). *Let,*

$$\delta = \begin{cases} d_1/\sigma, & \text{if } \{u_t\} \text{ is } I(0) \\ d_1/d(1), & \text{if } \{u_t\} \text{ is } I(1) \end{cases}.$$

Then, under the local alternative as defined by (33) − (35), *as* $T \to \infty$,

$$t_1 \Rightarrow \tau_{1,\infty} = \begin{cases} \left(\dfrac{\delta}{2} - 3\left[\frac{1}{3}W(1) - \int_0^1 W(r)\,dr\right]\right) \Big/ \sqrt{\phi(b,k)}, & \text{if } \{u_t\} \text{ is } I(0) \\ \left(\dfrac{\delta}{2} + 3\left[\frac{1}{3}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]\right) \Big/ \sqrt{\phi(b,k)}, & \text{if } \{u_t\} \text{ is } I(1) \end{cases}$$

and,

$$t_2 \Rightarrow \tau_{2,\infty} = \begin{cases} \left(-\dfrac{\delta}{\sqrt{12}} + \sqrt{12}\left[\frac{1}{2}W(1) - \int_0^1 W(r)\,dr\right]\right) \Big/ \sqrt{\phi(b,k)}, & \textit{if } \{u_t\} \textit{ is } I(0) \\ \left(-\dfrac{\delta}{\sqrt{12}} - \sqrt{12}\left[\frac{1}{2}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]\right) \Big/ \sqrt{\phi(b,k)}, & \textit{if } \{u_t\} \textit{ is } I(1) \end{cases}$$

and,

$$PLR_{UR} \Rightarrow 1(\tau_{1,\infty}\tau_{2,\infty} < 0) \cdot \min\left[\tau_{1,\infty}^2, \tau_{2,\infty}^2\right] \cdot \exp\left(-c_{UR}UR_{\overline{\alpha}}^{\infty}\right).$$

By means of simulation methods, similar to those that were used to generate asymptotic null distributions, the limiting distributions of PLR_{UR} tests reported above were computed under local alternatives for various values of δ. When computing asymptotic power, rejection probabilities were obtained by using 10% asymptotic null critical values. For each kernel, the power analysis is carried out for the grid of bandwidths given by $b = 0.02, 0.04, \ldots, 1.0$.

When analyzing asymptotic power of PLR_{UR}, stationary and unit root local designs need to be examined separately, because asymptotic distributions of tests are different in each of the two cases. When the errors are stationary, the scaling factors converge to zero asymptotically, and hence do not play an asymptotic role in determining power across kernels and bandwidths. However, for the case of unit root errors, the asymptotic power also depends on whether J or BG statistic is used in forming the scaling factor.

For both stationary and unit root cases, it is possible to define and plot power envelopes such that, for each value of δ, the point on the power envelope is the maximum attainable power across all five kernels and the grid of bandwidths (and also across the two scaling factors for the unit root case). Then, the asymptotic power performance of the PLR_{UR} test with a specific kernel, bandwidth and scaling factor can be evaluated by the closeness to the power envelope.

In Figure 2, asymptotic power is plotted for the case of stationary errors. The power envelope is plotted along with power obtained when each of five kernels is used with the smallest possible bandwidth in the grid of bandwidths considered, which is $b = 0.02$. It is clear from Figure 2 that regardless of the choice of kernel, the power envelope is attained whenever $b = 0.02$. Therefore, when the errors are stationary, using any of the five kernels with $b = 0.02$ should deliver essentially asymptotically power optimal tests within the class of tests considered here.

Figures 3 and 4 illustrate asymptotic power of tests when the errors exhibit a pure unit root process $(\overline{\alpha} = 0)$. Both figures plot the overall power envelope for the pure unit root case. Figures 3 and 4 also plot power envelopes conditional on kernel and scaling factors. Both figures use the same axes and plot the same overall power envelope, and therefore, comparisons between the two are easy. It is clear that tests that use the Daniell kernel with BG scaling factor attain the overall power envelope when $\overline{\alpha} = 0$.

Figure 5, on the other hand, plots the overall power envelope for $\overline{\alpha} = 0$, and also plots power for specific bandwidth values when Daniell kernel is used along with BG scaling factor. This figure illustrates that tests that use the Daniell kernel with $b = 0.22$ and BG scaling factor attain the power envelope when errors have a pure unit root, and therefore, such tests are asymptotically power optimal within the class of tests considered here.

Based on Figures 3-5, it is recommended that the Daniell kernel be used in practice with $b = 0.22$ and BG scaling factor, when errors have a pure unit root. When errors are stationary, the choice of kernel does not matter (as illustrated by Figure 2), but for the sake of convenience, it is recommended that the Daniell kernel be used with $b = 0.02$ and BG scaling factor.[6]

One concern to practitioners might be that when the Daniell kernel is implemented with $b = 0.02$ and BG scaling factor, the asymptotic power curve attained could be far from the power envelope curve, if, for example, errors are local to a unit root (see, Figures 7 and 8). This is relevant empirically as well, because, in a sample of size 100, $\overline{\alpha} = 10, 20$ correspond to AR(1) processes with AR(1) coefficients of 0.9 and 0.8 respectively. Another concern in practice might be that when the Daniell kernel is implemented with $b = 0.22$ and BG scaling factor to suit an underlying unit root local error structure, it turns out that errors are stationary. Then, asymptotic power curve attained will be sub-optimal, as illustrated in Figure 6.

Suppose that the Daniell kernel is implemented with $b = 0.12$ and BG scaling factor. This choice of bandwidth delivers tests with asymptotic power close to the optimal envelope when errors are stationary (see, Figure 6) or have a pure unit root (see, Figure 5). However, when errors are local to a unit root (see, Figures 7 and 8), Daniell kernel with $b = 0.12$ and BG scaling factor delivers tests that are more powerful than if $b = 0.02$ or $b = 0.22$ were used. Therefore, if the slight loss in asymptotic power in stationary and pure unit root cases when using $b = 0.12$ is a small enough price to pay for having more power in unit root local cases, for "insuring" against the above stated concerns, and most importantly for convenience, then it is recommended that $b = 0.12$ is used in practice with the Daniell kernel and BG scaling factor, at all times. Indeed, in the next section, it will be illustrated that $b = 0.12$ can deliver higher finite sample power than by $b = 0.02$ or $b = 0.22$, when errors are stationary.

6. Finite Sample Analysis

In this section, finite sample size and power performance of PLR_{UR} is examined when the recommended kernel and bandwidths are used. This is achieved by means of Monte Carlo simulation

[6]The choice of scaling factor does not matter asymptotically when errors are stationary. However, when errors are unit root local, i.e. $\overline{\alpha} = 10, 20$, the Daniell kernel with $b = 0.02$ was found to deliver higher power when used along with BG scaling factor rather than J scaling factor.

methods.

The data is generated according to model (1) with $ARMA(1,1)$ errors, as described in (25). For each test, null hypothesis rejection probabilities are calculated for various values of ρ and θ by using 10% asymptotic critical values from Table 2. Empirical rejection probabilities are reported in Tables 3-5 for sample sizes equal to 50, 100, and 200. 10,000 replications were performed in each case.

It is evident in Tables 3-5 that all recommended bandwidths for the Daniell kernel, along with BG scaling factor, deliver tests with empirical rejection probabilities that are either close to 0.10 or lower, except only when a large negative MA term is present together with a large positive AR term. This clearly indicates that scaling factor approach works well in practice when compared to the conventional HAC tests which usually suffer from severe over-rejection problems.

The reason for overrejection when a large negative MA term is present together with a large positive AR term is because BG unit root statistic is oversized for testing the unit root null in finite samples, and therefore, PLR_{BG} is not scaled down enough to eliminate overrejection completely.

Finite sample power simulation results are plotted in Figures 9-20. The power is not size adjusted, because the tests are size-robust by design, allowing more meaningful comparisons of actual power obtained when tests are implemented. It is interesting to note that even though $b = 0.02$ delivers asymptotically optimal power when errors are stationary, this is not the case in finite samples. This is because, the scaling factors matter in finite samples even when $\{u_t\}$ is i.i.d., delivering undersized tests when $b = 0.02$.

Figures 9 and 10 illustrate that, when errors are stationary and sample size is equal to 50, $b = 0.12$ delivers higher power than that obtained by using $b = 0.02$ or $b = 0.22$. As the sample size gets larger, however, tests using Daniell-BG with $b = 0.02$ start to dominate, as expected (see Figures 13, 14, 17 and 18). When errors are local to a unit root, the power obtained is qualitatively similar to local asymptotic power: Daniell-BG with $b = 0.22$ delivers the highest power, followed closely by Daniell-BG with $b = 0.12$ (see Figures 11, 12, 15, 16, 19, and 20).

7. Evidence on U.S. Regional Convergence

This section provides empirical evidence on U.S. regional convergence. Using statistics developed in the paper, β-convergence is tested among eight regions of the United States: New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountains, and Far West. These regions are defined by the BEA, and a list of states constituting each region can be found in Table 6. Carlino and Mills (1993), Loewy and Papell (1996), and Tomljanovic and Vogelsang (2002) report the existence of a break in trend at the end of WWII, consequently, the data analyzed here cover the period 1947-2002. The data are obtained from the BEA, and consists of the time-series of annual personal incomes per capita for each of the eight regions.[7] For each time-series, the natural logarithm of the ratio of per-capita income to cross-regional average per-capita income is computed. In the notation of Section 3, the computed series correspond to $\{y_t\}$ for each of the eight regions. The log-ratio series are plotted in Figures 21-29.

The model in (1) is estimated by OLS for each of the eight regions. Estimation results are reported in Table 7 along with test statistics that were recommended for practical applications earlier in the paper. For testing equality to zero of each of the estimated regression intercept and slope parameters, serial correlation robust and powerful two-sided t-tests, t_J and t_{BG}, are also reported in Table 7. t_J and t_{BG} tests have been recently proposed by Bunzel and Vogelsang (2003). For power optimality, they recommend using t_{BG} with Daniell kernel and $b = 0.16$ in the case of unit root errors, and t_J with Daniell kernel and $b = 0.02$ in the case of stationary errors. Furthermore, they show that the tests also have good finite sample size and power properties under a variety of ARMA$(1,1)$ error specifications.

Using the recommended t_J and t_{BG} tests, none of the estimated slope parameters are found to be statistically different from zero at 10% significance level, except for the Great Lakes region. It is important to note that t_J and t_{BG} tests are robust to highly persistent serial correlation, and a unit root in the errors, because, indeed, the null hypothesis of a unit root in the errors cannot be rejected at 10% level for any of the regions, except for the Plains.

Using the PLR test, the null hypothesis as defined in (2)-(4) is rejected for Great Lakes, Southeast and Farwest. Consequently, for these regions, β-convergence occurrence is verified. Figures 23, 25, and 28 also support this finding. On the other hand, in the remaining regions where the null hypothesis is not rejected, namely, New England, Mideast, Plains, Southwest and Rocky Mountains, no evidence in support of convergence is found. This result is well-supported by Figures 21, 22, 24, 26 and 27 as well.

8. Conclusion

In this paper, a likelihood ratio approach is taken to derive a test of the economic convergence hypothesis in the context of the linear deterministic trend model. The test is designed to directly address the nonstandard nature of the hypothesis, and is a systematic improvement over existing methods for testing convergence in the same context. The test is first derived under the assumption of Gaussian errors with known serial correlation. However, the normality assumption is then relaxed, and the results are naturally extended to the case of covariance stationary errors with unknown serial correlation. The test statistic is a continuous function of the individual t-statistics

[7] Ideally, the incomes should be deflated by using regional price deflators, however, regional price deflation indexes are not available for the regions under consideration.

on the intercept and slope parameters of the linear trend model. This allows direct implementation of standard heteroskedasticity and autocorrelation consistent estimators of the long-run variance.

Building upon the likelihood ratio framework, concrete and specific tests are recommended to be used in practice. The recommended tests do not require the knowledge of the form of serial correlation in the data, and they are robust to highly persistent serial correlation and a unit root in the errors. Furthermore, the tests have asymptotic power close to that of the known variance case. The recommended tests utilize the nonparametric kernel variance estimators, based on the fixed bandwidth (fixed-b) asymptotic framework recently proposed by Kiefer and Vogelsang (2003). The fixed-b framework makes possible the choice of kernel and bandwidth that deliver tests with maximal asymptotic power within a specific class of tests.

It is shown that when the Daniell kernel variance estimator is implemented with specific bandwidth choices, the recommended tests have asymptotic power close that of the known variance case, as well as good finite sample size and power properties. In particular, when the errors are stationary, the Breitung scaling factor and the Daniell kernel with bandwidth parameter equal to 0.02 times the sample size provide a test with optimal power. On the other hand, when errors have a unit root, the Breitung scaling factor and the Daniell kernel with bandwidth parameter equal to 0.22 times the sample size deliver a power optimal test. For convenience, the bandwidth parameter equal to 0.12 times the sample size is also recommended to be used with Breitung scaling factor and the Daniell kernel to obtain generally high power in both stationary and unit root cases.

Finally, the recommended tests are used to investigate economic convergence among eight regions of the US, as defined by the Bureau of Economic Analysis, in the post WWII period. Evidence for convergence is found in three of the eight regions.

References

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–854.

Barro, R. J. (1991): "Economic Growth in a Cross Section of Countries," *The Quarterly Journal of Economics*, 106, 407–443.

Barro, R. J., and X. Sala-i-Martin (1992): "Convergence," *Journal of Political Economy*, 100, 223–251.

——— (1995): "Technological Diffusion, Convergence and Growth," NBER Working Paper 5151.

Baumol, W. (1986): "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show," *American Economic Review*, 76, 1072–1085.

Baumol, W. J., R. R. Nelson, and E. N. Wolff(eds) (1994): *Convergence of Productivity: Cross National Studies and Historical Evidence*. Oxford University Press, New York.

Bernard, A. B., and S. N. Durlauf (1991): "Convergence of International Output Movements," NBER Working Paper, No. 3717.

——— (1995): "Convergence in International Output," *Journal of Applied Econometrics*, 10, 161–173.

——— (1996): "Interpreting Tests for the Convergence Hypothesis," *Journal of Econometrics*, 91, 161–173.

Breitung, J. (2002): "Nonparametric Tests for Unit Roots and Cointegration," *Journal of Econometrics*, 108, 343–363.

Brown, S. J., N. E. Coulson, and R. F. Engle (1990): "Non-cointegration and Econometric Evaluation of Models of Regional Shift and Share," NBER Working Paper, No. 3291.

Bunzel, H., and T. J. Vogelsang (2003): "Powerful Trend Function Tests That are Robust to Strong Serial Correlation with an Application to the Prebish Singer Hypothesis," Cornell University, manuscript.

Carlino, G. A., and L. O. Mills (1993): "Are U.S. Regional Incomes Converging?," *Journal of Monetary Economics*, 32, 335–346.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, 25, 573–578.

den Haan, W., and A. Levin (1998): "Vector Autoregressive Covariance Matrix Estimation," University of California, San Diego, manuscript.

——— (2000): "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," NBER Working Paper, No. 255.

Feder, P. I. (1968): "On the Distribution of the Log Likelihood Ratio Test Statistic When the True Parameter is 'Near' the Boundaries of the Hypothesis Regions," *The Annals of Mathematical Statistics*, 39, 2044–2055.

Friedman, M. (1992): "Do old fallacies never die?," *Journal of Economic Literature*, 30, 2129–2132.

Gouriéroux, C., A. Holly, and A. Monfort (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models With Inequality Constraints on the Regression Parameters," *Econometrica*, 50, 63–80.

Grenander, U., and M. Rosenblatt (1957): *Statistical Analysis of Stationary Time Series*. Chelsea Publishing Company, New York.

Hobijn, B., and P. Franses (2000): "Asymptotically Perfect and Relative Convergence of Productivity," *Journal of Applied Econometrics*, 15, 59–81.

Kiefer, N. M., and T. J. Vogelsang (2003): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," Cornell University, manuscript.

Kudô, A. (1963): "A Multivariate Analogue of the One-Sided Test," *Biometrika*, 50, 403–418.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root: How Sure are we that Economic Time Series Have a Unit Root?," *Journal of Econometrics*, 54, 159–178.

Linden, M. (2000): "Testing Growth Convergence with Time Series Data - A Non-Parametric Approach," *International Review of Applied Economics*, 14, 361–370.

Loewy, M. B., and D. H. Papell (1996): "Are U.S. Regional Incomes Converging? Some Further Evidence," *Journal of Monetary Economics*, 38, 587–598.

Mankiw, N., D. Romer, and D. Weil (1992): "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 107, 407–437.

Park, J. Y. (1990): "Testing for Unit Roots and Cointegration by Variable Addition," *in Advances in Econometrics: Cointegration, Spurious Regressions and Unit Roots*, ed. by T. Fomby and F. Rhodes. London: Jai Press, pp. 107-134.

Park, J. Y., and B. Choi (1988): “A New Approach to Testing for a Unit Root,” CAE Working Paper, No. 88-23, Cornell University.

Perlman, M. D. (1969): “One-Sided Testing Problems in Multivariate Analysis,” *The Annals of Mathematical Statistics*, 40, 549–567.

Phillips, P. C. B. (1987): “Time Series Regression with Unit Roots,” *Econometrica*, 55, 277–302.

Phillips, P. C. B., and V. Solo (1992): “Asymptotics for Linear Processes,” *The Annals of Statistics*, 20, 971–1001.

Quah, D. (1993): “Galton’s Fallacy and Tests of the Convergence Hypothesis,” *Scandinavian Journal of Economics*, 93, 427–443.

——— (1996): “Twin peaks: Growth and Convergence in Models of Distributional Dynamics,” *The Economic Journal*, 106, 1045–1055.

Schwarz, G. (1978): “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6, 461–464.

Self, S. G., and K. Y. Liang (1987): “Asymptotic Properties of Maximum Likelihood Ratio Tests Under Nonstandard Conditions,” *Journal of the American Statistical Association*, 82, 605–610.

Solow, R. M. (1956): “A Contribution in the Theory of Economic Growth,” *The Quarterly Journal of Economics*, 70, 65–94.

Tanaka, K. (1990): “Testing for a Moving Average Root,” *Econometric Theory*, 6, 433–444.

Tomljanovic, M., and T. J. Vogelsang (2002): “Are U.S. Regions Converging? Using New Econometric Methods to Examine Old Issues,” *Empirical Economics*, 27, 49–62.

Vogelsang, T. J. (1998): “Trend Function Hypothesis Testing in the Presence of Serial Correlation,” *Econometrica*, 65, 123–148.

Wolak, F. A. (1989): “Testing Inequality Constraints in Linear Econometric Models,” *Journal of Econometrics*, 41, 205–235.

Yancey, T. A., G. G. Judge, and M. E. Bock (1981): “Testing Multiple Equality and Inequality Hypothesis in Economics,” *Economic Letters*, 7, 249–255.

Appendix

A.1. Supplementary Results

The following supplementary results are fairly standard (see, for example, Vogelsang (1998) and Bunzel and Vogelsang (2003)). They are stated without further proof.

Lemma 3 *Suppose that* $\{u_t\}$ *is* $I(0)$. *Let* $\boldsymbol{\tau}_T \equiv \left[\begin{smallmatrix} 1 & 0 \\ 0 & T^{-1} \end{smallmatrix}\right]$. *Then, as* $T \to \infty$,

$$(a) \quad T^{1/2}\boldsymbol{\tau}_T^{-1}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \Longrightarrow \sigma \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} W(1) \\ W(1) - \int_0^1 W(r)\,dr \end{bmatrix}, \tag{36}$$

$$(b) \quad T^{-1/2}\widehat{S}_{[rT]} \Longrightarrow \sigma \widehat{Q}_0(r), \tag{37}$$

where $\widehat{Q}_0(r) \equiv W(r) + r(2-3r)W(1) + 6r(r-1)\int_0^1 W(s)\,ds$.

Lemma 4 *Suppose that* $\{u_t\}$ *is* $I(1)$. *Then, as* $T \to \infty$,

$$(a) \quad T^{-1/2}\boldsymbol{\tau}_T^{-1}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \Longrightarrow d(1) \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix} \begin{bmatrix} \int_0^1 V_{\overline{\alpha}}(r)\,dr \\ \int_0^1 r V_{\overline{\alpha}}(r)\,dr \end{bmatrix}, \tag{38}$$

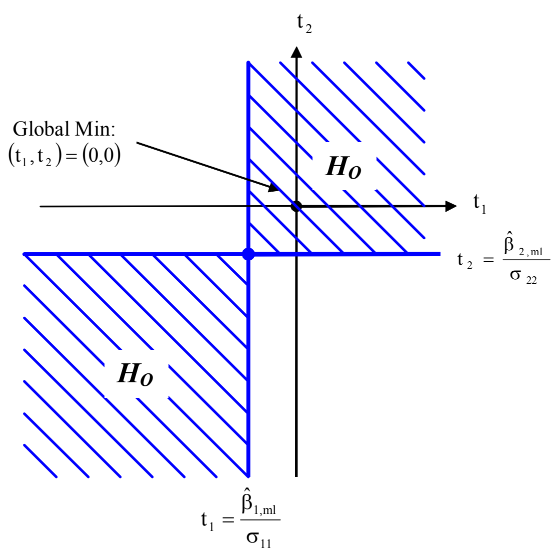
$$(b) \quad T^{-1/2}\widehat{u}_{[rT]} \Longrightarrow d(1)\widehat{Q}_1(r),$$

$$(c) \quad T^{-3/2}\widehat{S}_{[rT]} \Longrightarrow d(1)\int_0^r \widehat{Q}_1(s)\,ds, \tag{39}$$

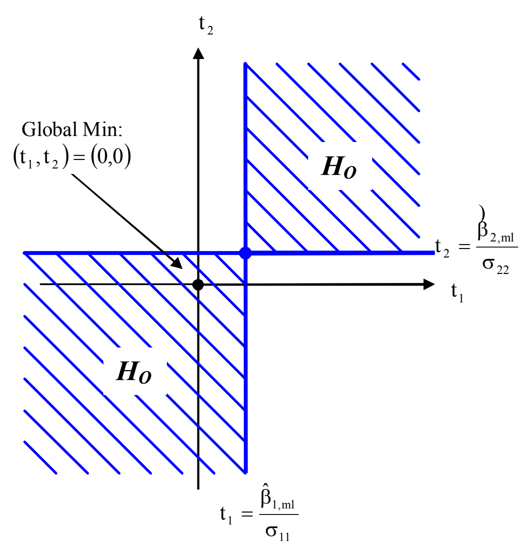
where, $\widehat{Q}_1(s) \equiv V_{\overline{\alpha}}(s) - (4-6s)\int_0^1 V_{\overline{\alpha}}(r)\,dr + (6-12s)\int_0^1 rV_{\overline{\alpha}}(r)\,dr$.

A.2. Proofs

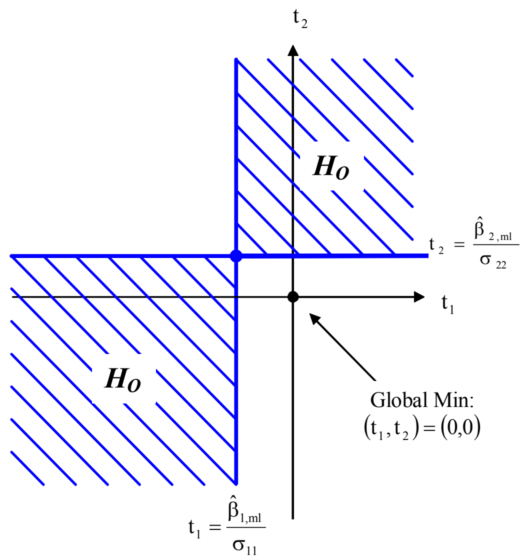
Proof of Theorem 2: According to the signs of $\widehat{\boldsymbol{\beta}}_{1,ml}\big/\sigma_{11}$ and $\widehat{\boldsymbol{\beta}}_{2,ml}\big/\sigma_{22}$, there are four cases to be considered in evaluating (16) :



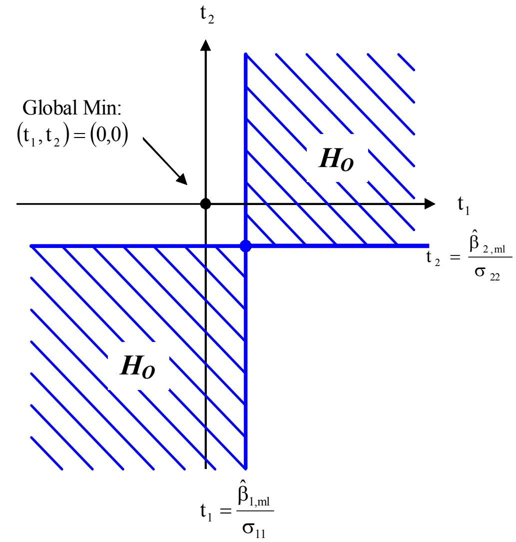
Case 1



Case 2



Case 3



Case 4

In each of the four cases, (t_1, t_2) that minimizes $t_1^2 + t_2^2 - 2\rho_{12}t_1t_2$ under the null hypothesis can be chosen in the traced areas inclusive of the boundaries. In cases 1 and 2, the global minimum, $(t_1, t_2) = (0, 0)$, is included under the null hypothesis and it is thus chosen as the minimizing solution. In cases 3 and 4, the minimizer is found along either of the boundaries, $t_1 = \widehat{\beta}_{1,ml} \big/ \sigma_{11}$ or $t_2 = \widehat{\beta}_{2,ml} \big/ \sigma_{22}$. Let $f(t_1, t_2) = \left(1 - \rho_{12}^2\right)^{-1} \left(t_1^2 + t_2^2 - 2\rho_{12}t_1t_2\right)$. Then,

$$\left.\frac{\partial f(t_1, t_2)}{\partial t_1}\right|_{t_1 = \frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}} = 0 \rightarrowtail t_2 = \rho_{12}\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}} \rightarrowtail f\left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}, \rho_{12}\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\right) = \left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\right)^2, \tag{40}$$

and

$$\left.\frac{\partial f(t_1, t_2)}{\partial t_2}\right|_{t_1 = \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}} = 0 \rightarrowtail t_1 = \rho_{12}\frac{\widehat{\beta}_{2,ml}}{\sigma_{22}} \rightarrowtail f\left(\rho_{12}\frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}, \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\right) = \left(\frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\right)^2. \tag{41}$$

Consequently, the argument that solves (16) is,

$$\arg \min_{(t_1,t_2)\in\boldsymbol{\omega}_0'} t_1^2 + t_2^2 - 2\rho_{12}t_1t_2 =$$

$$\begin{cases} \left\{\begin{matrix} \left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\ ,\ \rho_{12}\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\right) \\ OR \\ \left(\rho_{12}\frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\ ,\ \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\right) \end{matrix}\right\}, & if \left\{\begin{matrix} \left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}} > 0\ ,\ \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}} < 0\right) \\ OR \\ \left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}} < 0\ ,\ \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}} > 0\right) \end{matrix}\right. \\ (0,0)\,, & otherwise \end{cases} . \tag{42}$$

The test statistic is then given by,

$$LR = \begin{cases} \min\left[\left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}}\right)^2, \left(\frac{\widehat{\beta}_{2,ml}}{\sigma_{22}}\right)^2\right], & if \begin{cases} \left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}} > 0\ ,\ \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}} < 0\right) \\ OR \\ \left(\frac{\widehat{\beta}_{1,ml}}{\sigma_{11}} < 0\ ,\ \frac{\widehat{\beta}_{2,ml}}{\sigma_{22}} > 0\right) \end{cases} \\ \\ 0, & otherwise \end{cases} . \tag{43}$$

Note that, if $t_i = \left(\widehat{\beta}_{i,ml} - \beta_i\right) \Big/ \sigma_{ii}$ for $i = 1, 2$, are evaluated at the dividing boundary of the null hypothesis, $\beta_1 = \beta_2 = 0$, then (43) can simply be written as

$$LR = 1\left(t_1 \cdot t_2 < 0\right) \cdot \min\left[t_1^2\,,\, t_2^2\right], \tag{44}$$

where $1(\cdot)$ is the indicator function given by

$$1\left(t < 0\right) = \begin{cases} 1, & if\ t < 0 \\ 0, & otherwise \end{cases} .$$

■

Proof of Theorem 3: Let the cumulative distribution function of $PLR = 1\left(t_1 t_2 < 0\right) \min\left[t_1^2, t_2^2\right]$ be denoted by $F_{PLR}(v)$. Denote $Y\left(t_1, t_2\right) = 1\left(t_1 t_2 < 0\right)$ and $Z\left(t_1, t_2\right) = \min\left[t_1^2, t_2^2\right]$. Then, by definition of c.d.f.,

$$\begin{aligned} F_{PLR}(v) &= P\left[PLR \le v\right] \\ &= P\left[Y \cdot Z \le v\right] \\ &= P\left[\left(Y \cdot Z \le v\right) \cap \left(Y = 1\right)\right] + P\left[\left(Y \cdot Z \le v\right) \cap \left(Y = 0\right)\right]. \end{aligned} \tag{45}$$

Considering each term in (45) separately, it is obtained that,

$$\begin{aligned} P\left[\left(Y \cdot Z \le v\right) \cap \left(Y = 1\right)\right] &= P\left[Y \cdot Z \le v \mid Y = 1\right] \cdot P\left[Y = 1\right] \\ &= P\left[Z \le v\right] \cdot P\left[Y = 1\right], \end{aligned}$$

and,

$$\begin{aligned} P\left[\left(Y \cdot Z \le v\right) \cap \left(Y = 0\right)\right] &= P\left[Y \cdot Z \le v \mid Y = 0\right] \cdot P\left[Y = 0\right] \\ &= P\left[0 \le v\right] \cdot P\left[Y = 0\right]. \end{aligned}$$

Let $p = P\left[Y = 1\right]$ Then, $1 - p = P\left[Y = 0\right]$, and (45) can be written as,

$$F_{PLR}(v) = p \cdot F_Z(v) + (1 - p) \cdot \theta(v),$$

where,

$$\theta(v) = \begin{cases} 0, & \text{if } v < 0 \\ 1, & \text{if } v \ge 0 \end{cases} .$$

Then, by definition of p.d.f.,

$$f_{PLR}(v) = \frac{dF_{PLR}(v)}{dv} = p \cdot f_Z(v) + (1-p)\,\delta(v), \tag{46}$$

where $\delta(v)$ is the dirac delta function at zero, defined by $\delta(v) = 0$ if $v \neq 0$, and $\int_{-v_1}^{v_2} \delta(v)\ dv = 1$ for $v_1 > 0$ and $v_2 > 0$.

The value of p in (46) can be easily obtained by using the bivariate normal distribution of t_1 and t_2. To obtain $f_Z(v)$, start with the definition of c.d.f. of Z and write,

$$\begin{aligned} F_Z(v) &= P[Z \leq v] = P\left[\min\left[t_1^2, t_2^2\right] \leq v\right] \\ &= P\left[\left(t_1^2 \leq v\right) \cap \left(t_1^2 < t_2^2\right)\right] + P\left[\left(t_2^2 \leq v\right) \cap \left(t_2^2 \leq t_1^2\right)\right] \\ &= P\left[\left(-\sqrt{v} \leq t_1 \leq \sqrt{v}\right) \cap \left((t_1 - t_2)(t_1 + t_2) < 0\right)\right] \\ &\qquad + P\left[\left(-\sqrt{v} \leq t_2 \leq \sqrt{v}\right) \cap \left((t_1 - t_2)(t_1 + t_2) \geq 0\right)\right] \\ &= 1 - 2\iint\limits_A f_{t_1 t_2}(t_1, t_2)\ dt_1\ dt_2 - 2\iint\limits_B f_{t_1 t_2}(t_1, t_2)\ dt_1\ dt_2, \end{aligned} \tag{47}$$

where, the last equality follows by the symmetry of the joint p.d.f. of t_1 and t_2, given by $f_{t_1 t_2}(t_1, t_2)$, and,

$$\begin{aligned} A &= \left\{(t_1, t_2) : t_1 \in \left(\sqrt{v}, +\infty\right),\quad t_2\left(\sqrt{v}, +\infty\right)\right\}, \\ B &= \left\{(t_1, t_2) : t_1 \in \left(\sqrt{v}, +\infty\right),\quad t_2\left(-\infty, -\sqrt{v}\right)\right\}. \end{aligned}$$

By using the formula, $f_{t_1 t_2}(t_1, t_2) = f_{t_2|t_1}(t_2 \mid t_1) \cdot f_{t_1}(t_1)$, integrals in (47) can be written as,

$$\iint\limits_A f_{t_1 t_2}(t_1, t_2)\ dt_1\ dt_2 = \int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1 \int_{\sqrt{v}}^{\infty} f_{t_2|t_1}(t_2 \mid t_1)\ dt_2, \tag{48}$$

$$\iint\limits_B f_{t_1 t_2}(t_1, t_2)\ dt_1\ dt_2 = \int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1 \int_{-\infty}^{-\sqrt{v}} f_{t_2|t_1}(t_2 \mid t_1)\ dt_2. \tag{49}$$

Note that, since t_1 and t_2 are bivariate normal (see, (21)), it is well-known that,

$$(t_2 \mid t_1) \sim \mathbb{N}\left(-\frac{\sqrt{3}}{2}t_1, 1 - \left(-\frac{\sqrt{3}}{2}\right)^2\right) = \mathbb{N}\left(-\frac{\sqrt{3}}{2}t_1, \frac{1}{4}\right),$$

and therefore,

$$\begin{aligned} \int_{\sqrt{v}}^{\infty} f_{t_2|t_1}(t_2 \mid t_1)\ dt_2 &= P\left[W_{\mathbb{N}\left(-\frac{\sqrt{3}}{2}t_1, \frac{1}{4}\right)} > \sqrt{v}\right] \\ &= 1 - P\left[W_{\mathbb{N}(0,1)} \leq \frac{\sqrt{v} + \frac{\sqrt{3}}{2}t_1}{\sqrt{\frac{1}{4}}}\right] \\ &= 1 - \phi\left(2\sqrt{v} + \sqrt{3}t_1\right), \end{aligned} \tag{50}$$

where $\phi(\cdot)$ is the standard normal c.d.f.. Similarly,

$$\int_{-\infty}^{-\sqrt{v}} f_{t_2|t_1}(t_2 \mid t_1)\ dt_2 = P\left[W_{\mathbb{N}\left(-\frac{\sqrt{3}}{2}t_1,\frac{1}{4}\right)} \leq -\sqrt{v}\right] = \phi\left(-2\sqrt{v}+\sqrt{3}t_1\right). \tag{51}$$

Substituting (50) and (51) into (48) and (49), and then, (48) and (49) back into (47), it is obtained that,

$$\begin{aligned} F_Z(v) &= 1-2\left(\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1 - \int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,\phi\left(2\sqrt{v}+\sqrt{3}t_1\right)\ dt_1\right) \\ &\qquad -2\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1\ \phi\left(-2\sqrt{v}+\sqrt{3}t_1\right)\ dt_1 \\ &= 1-2\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1 + 2\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,\phi\left(2\sqrt{v}+\sqrt{3}t_1\right)\ dt_1 \\ &\qquad -2\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1\ \phi\left(-2\sqrt{v}+\sqrt{3}t_1\right)\ dt_1. \end{aligned} \tag{52}$$

Note that $f_{t_1}(\cdot)$ is the marginal p.d.f. of t_1, which is $\mathbb{N}(0,1)$. Then, by definition of p.d.f.,

$$\begin{aligned} f_Z(v) &= \frac{d}{dv}F_Z(v) \\ &= -2\frac{d}{dv}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1 + 2\frac{d}{dv}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,\phi\left(2\sqrt{v}+\sqrt{3}t_1\right)dt_1 \\ &\qquad -2\frac{d}{dv}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,\phi\left(-2\sqrt{v}+\sqrt{3}t_1\right)\ dt_1. \end{aligned} \tag{53}$$

By using the First Fundamental Theorem of Integral Calculus and evaluating term by term, the first term in (53) is given by,

$$-2\frac{d}{dv}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\ dt_1 = -\frac{1}{\sqrt{v}}f_{t_1}\left(\sqrt{v}\right), \tag{54}$$

the second term in (53) is given by,

$$\begin{aligned} 2\frac{d}{dv}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,\phi\left(2\sqrt{v}+\sqrt{3}t_1\right)dt_1 &= -\frac{1}{\sqrt{v}}f_{t_1}\left(\sqrt{v}\right)\phi\left(2\sqrt{v}+\sqrt{3}\sqrt{v}\right) \\ &\qquad +\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,f\left(2\sqrt{v}+\sqrt{3}t_1\right)dt_1, \end{aligned} \tag{55}$$

and, the third term in (53) is given by,

$$\begin{aligned} -2\frac{d}{dv}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,\phi\left(-2\sqrt{v}+\sqrt{3}t_1\right)\ dt_1 &= \frac{1}{\sqrt{v}}f_{t_1}\left(\sqrt{v}\right)\phi\left(-2\sqrt{v}+\sqrt{3}\sqrt{v}\right) \\ &\qquad +\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f_{t_1}(t_1)\,f\left(-2\sqrt{v}+\sqrt{3}t_1\right)dt_1, \end{aligned} \tag{56}$$

where $f(\cdot)$ is the standard normal p.d.f.. For consistency, replace $f_{t_1}(\cdot)$ with $f(\cdot)$.

The second term in (55) is,

$$\begin{aligned}\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f(t)\, f\left(2\sqrt{v}+\sqrt{3}t\right) dt &= \frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\frac{1}{\sqrt{2\pi}}e^{-\left(2\sqrt{v}+\sqrt{3}t\right)^2/2}dt \\ &= \frac{1}{2\pi}\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} e^{-\frac{1}{2}\left(t^2+\left(2\sqrt{v}+\sqrt{3}t\right)^2\right)}dt. \qquad (57)\end{aligned}$$

Rearranging the exponential function in (57),

$$\begin{aligned}t^2+\left(2\sqrt{v}+\sqrt{3}t\right)^2 &= t^2+4v+4\sqrt{3}\sqrt{v}t+3t^2 = 4t^2+4v+4\sqrt{3}\sqrt{v}t \\ &= 4\left(t+\frac{\sqrt{3}}{2}\sqrt{v}\right)^2+v,\end{aligned}$$

and plugging it back into (57),

$$\begin{aligned}\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f(t)\, f\left(2\sqrt{v}+\sqrt{3}t\right) dt &= \frac{1}{2\pi}\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} e^{-\frac{1}{2}\left(4\left(t+\frac{\sqrt{3}}{2}\sqrt{v}\right)^2+v\right)}dt \\ &= \frac{1}{2\pi}\frac{2}{\sqrt{v}}e^{-v/2}\int_{\sqrt{v}}^{\infty} e^{-\frac{4}{2}\left(t+\frac{\sqrt{3}}{2}\sqrt{v}\right)^2}dt\end{aligned}$$

Make a change of variable by letting $u=2\left(t+\frac{\sqrt{3}}{2}\sqrt{v}\right)$. Then, $\frac{u}{2}-\frac{\sqrt{3}}{2}\sqrt{v}=t$, and, $dt=\frac{du}{2}$. Also, $t=\sqrt{v}$ implies $u=2\left(1+\frac{\sqrt{3}}{2}\right)\sqrt{v}=\left(2+\sqrt{3}\right)\sqrt{v}$. Then, the second term in (55) becomes,

$$\begin{aligned}\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f(t)\, f\left(2\sqrt{v}+\sqrt{3}t\right) dt &= \frac{1}{2\pi}\frac{2}{\sqrt{v}}e^{-v/2}\int_{\left(2+\sqrt{3}\right)\sqrt{v}}^{\infty} e^{-u^2/2}\,\frac{du}{2} \\ &= \frac{1}{\sqrt{v}}\frac{1}{\sqrt{2\pi}}e^{-v/2}\frac{1}{\sqrt{2\pi}}\int_{\left(2+\sqrt{3}\right)\sqrt{v}}^{\infty} e^{-u^2/2}du \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}}e^{-v/2}\left[1-\phi\left(\left(2+\sqrt{3}\right)\sqrt{v}\right)\right]. \qquad (58)\end{aligned}$$

Similarly, the second term in (56) can be written as,

$$\frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f(t)\, f\left(-2\sqrt{v}+\sqrt{3}t\right) dt = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}}e^{-v/2}\left[1-\phi\left(\left(2-\sqrt{3}\right)\sqrt{v}\right)\right]. \qquad (59)$$

Finally, using (54)-(59), the equation in (53) becomes,

$$\begin{aligned}
f_Z(v) &= \frac{-1}{\sqrt{v}} f\left(\sqrt{v}\right) - \frac{1}{\sqrt{v}} f\left(\sqrt{v}\right)\phi\left(2\sqrt{v}+\sqrt{3}\sqrt{v}\right) + \frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f(t_1) f\left(2\sqrt{v}+\sqrt{3}t_1\right) dt_1 \\
&\quad + \frac{1}{\sqrt{v}} f\left(\sqrt{v}\right)\phi\left(-2\sqrt{v}+\sqrt{3}\sqrt{v}\right) + \frac{2}{\sqrt{v}}\int_{\sqrt{v}}^{\infty} f(t_1) f\left(-2\sqrt{v}+\sqrt{3}t_1\right) dt_1 \\
&= \frac{1}{\sqrt{v}} f\left(\sqrt{v}\right)\left[\phi\left(-2\sqrt{v}+\sqrt{3}\sqrt{v}\right) - \phi\left(2\sqrt{v}+\sqrt{3}\sqrt{v}\right) - 1\right] \\
&\quad + \frac{2}{\sqrt{v}}\left[\frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-v/2}\left[1-\phi\left(\left(2+\sqrt{3}\right)\sqrt{v}\right)\right] + \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-v/2}\left[1-\phi\left(\left(2-\sqrt{3}\right)\sqrt{v}\right)\right]\right] \\
&= \frac{1}{\sqrt{v}} f\left(\sqrt{v}\right)\left[\phi\left(-2\sqrt{v}+\sqrt{3}\sqrt{v}\right) - \phi\left(2\sqrt{v}+\sqrt{3}\sqrt{v}\right) - 1\right] \\
&\quad + \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}} e^{-v/2}\left[2 - \phi\left(2\sqrt{v}+\sqrt{3}\sqrt{v}\right) - \phi\left(2\sqrt{v}-\sqrt{3}\sqrt{v}\right)\right].
\end{aligned}$$

Note above that $\frac{1}{\sqrt{v}} f\left(\sqrt{v}\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}} e^{-v/2}$ is the chi-square p.d.f. with 1 degree of freedom. Also, note the symmetry of the standard normal c.d.f., which allows writing, $\phi\left(2\sqrt{v}-\sqrt{3}\sqrt{v}\right) = 1 - \phi\left(-2\sqrt{v}+\sqrt{3}\sqrt{v}\right)$. Then,

$$\begin{aligned}
f_Z(v) &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}} e^{-v/2}\left[\phi\left(\left(-2+\sqrt{3}\right)\sqrt{v}\right) - 2\phi\left(\left(2+\sqrt{3}\right)\sqrt{v}\right) - \phi\left(\left(2-\sqrt{3}\right)\sqrt{v}\right) + 1\right] \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}} e^{-v/2}\left[\phi\left(\left(-2+\sqrt{3}\right)\sqrt{v}\right) - 2\phi\left(\left(2+\sqrt{3}\right)\sqrt{v}\right) - 1 + \phi\left(\left(-2+\sqrt{3}\right)\sqrt{v}\right) + 1\right] \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}} e^{-v/2} 2\left[\phi\left(\left(-2+\sqrt{3}\right)\sqrt{v}\right) - \phi\left(\left(2+\sqrt{3}\right)\sqrt{v}\right)\right]. \qquad (60)
\end{aligned}$$

Note in (60) above that $\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v}} e^{-v/2}$ is the chi-square p.d.f. with 1 degree of freedom.

Finally, it is straightforward to show that $p = P\left[t_1 t_2 < 0\right] = \frac{5}{6}$ by using the bivariate normal distribution of t_1 and t_2. The result in the theorem then follows when (60) is used in (46).
■

Proof of Theorem 4: The results in the theorem follow by the continuous mapping theorem, Lemmas 3-4, and other results previously established in the paper. Note the following:
If $\{u_t\}$ is $I(0)$,

$$t_1 = \frac{\left(\widehat{\beta}_1 - \beta_1\right)}{\left[\widehat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{11}\right]^{1/2}} = \frac{T^{1/2}\left(\widehat{\beta}_1 - \beta_1\right)}{\left[\widehat{\sigma}^2 \frac{4T+2}{(T-1)}\right]^{1/2}} \Rightarrow \frac{-6\sigma\left[\frac{1}{3}W(1) - \int_0^1 W(r)\,dr\right]}{\left[4\sigma^2\phi(b,k)\right]^{1/2}},$$

$$t_2 = \frac{\left(\widehat{\beta}_2 - \beta_2\right)}{\left[\widehat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{22}\right]^{1/2}} = \frac{T^{3/2}\left(\widehat{\beta}_2 - \beta_2\right)}{\left[\widehat{\sigma}^2 \frac{12T^3}{T^3-T}\right]^{1/2}} \Rightarrow \frac{12\sigma\left[\frac{1}{2}W(1) - \int_0^1 W(r)\,dr\right]}{\left[12\sigma^2\phi(b,k)\right]^{1/2}},$$

and, if $\{u_t\}$ is $I(1)$,

$$t_1 = \frac{\left(\widehat{\beta}_1 - \beta_1\right)}{\left[\widehat{\sigma}^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{11}\right]^{1/2}} = \frac{T^{-1/2}\left(\widehat{\beta}_1 - \beta_1\right)}{\left[T^{-2}\widehat{\sigma}^2 \frac{4T+2}{(T-1)}\right]^{1/2}} \Rightarrow \frac{6d(1)\left[\frac{1}{3}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]}{\left[4d(1)^2\phi(b,k)\right]^{1/2}},$$

$$t_2 = \frac{\left(\widehat{\beta}_2 - \beta_2\right)}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{22}\right]^{1/2}} = \frac{T^{1/2}\left(\widehat{\beta}_2 - \beta_2\right)}{\left[T^{-2}\widehat{\sigma}^2 \frac{12T^3}{T^3-T}\right]^{1/2}} \Rightarrow \frac{-12d(1)\left[\frac{1}{2}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]}{\left[12d(1)^2\phi(b,k)\right]^{1/2}}.$$

■

Proof of Theorem 5: Let $\boldsymbol{\beta}_0 = [0\ \ 0]'$, and $\mathbf{D} = [d_1\ \ -d_1]'$. Recall that $\boldsymbol{\tau}_T = \left[\begin{smallmatrix}1 & 0\\ 0 & T^{-1}\end{smallmatrix}\right]$. Let

$$\mathbf{g}(T) = \begin{bmatrix} g_1(T) & 0 \\ 0 & g_2(T) \end{bmatrix},$$

where $g_1(T)$ and $g_2(T)$ are defined in (33). Note that $\mathbf{g}(T)$ satisfies,

$$\mathbf{g}(T) = \begin{cases} T^{-1/2}\boldsymbol{\tau}_T, & \text{if } \{u_t\} \text{ is } I(0) \\ T^{1/2}\boldsymbol{\tau}_T, & \text{if } \{u_t\} \text{ is } I(1) \end{cases}.$$

Then, under the local alternative,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \mathbf{g}(T)\mathbf{D} + \left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\sum_{t=1}^{T}\mathbf{x}_t u_t, \tag{61}$$

where, $\mathbf{x}_t = [1\ \ t]'$. Let $\mathbf{X}(r) = [1\ \ r]'$. It follows from (61) by Lemma 3, Lemma 4 and the continuous mapping theorem that, as $T \to \infty$, if $\{u_t\}$ is $I(0)$,

$$\begin{aligned} T^{1/2}\boldsymbol{\tau}_T^{-1}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) &= \begin{bmatrix} T^{1/2}\left(\widehat{\beta}_1 - 0\right) \\ T^{3/2}\left(\widehat{\beta}_2 - 0\right) \end{bmatrix} \\ &\Rightarrow \mathbf{D} + \sigma\left[\int_0^1 \mathbf{X}(r)\mathbf{X}(r)'\,dr\right]^{-1}\left[\int_0^1 \mathbf{X}(r)\,dW(r)\right] \\ &= \begin{bmatrix} d_1 \\ -d_1 \end{bmatrix} + \sigma\begin{bmatrix} -2W(1) + 6\int_0^1 W(r)\,dr \\ 6W(1) - 12\int_0^1 W(r)\,dr \end{bmatrix}, \end{aligned}$$

and, if $\{u_t\}$ is $I(1)$,

$$\begin{aligned} T^{-1/2}\boldsymbol{\tau}_T^{-1}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) &= \begin{bmatrix} T^{-1/2}\left(\widehat{\beta}_1 - 0\right) \\ T^{1/2}\left(\widehat{\beta}_2 - 0\right) \end{bmatrix} \\ &\Rightarrow \mathbf{D} + d(1)\left[\int_0^1 \mathbf{X}(r)\mathbf{X}(r)'\,dr\right]^{-1}\left[\int_0^1 \mathbf{X}(r)V_{\overline{\alpha}}(r)\,dr\right] \\ &= \begin{bmatrix} d_1 \\ -d_1 \end{bmatrix} + d(1)\begin{bmatrix} 4\int_0^1 V_{\overline{\alpha}}(r)\,dr - 6\int_0^1 rV_{\overline{\alpha}}(r)\,dr \\ -6\int_0^1 V_{\overline{\alpha}}(r)\,dr + 12\int_0^1 rV_{\overline{\alpha}}(r)\,dr \end{bmatrix}. \end{aligned}$$

Since the long-run variance estimator $\widehat{\sigma}^2$ and unit root statistics J and BG are invariant to the true value of β_1 and β_2, their limiting distributions are the same as under the null hypothesis. Then, it is obtained that, under the local alternative, as $T \to \infty$, if $\{u_t\}$ is $I(0)$,

$$\begin{aligned} t_1 &= \frac{\left(\widehat{\beta}_1 - 0\right)}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{11}\right]^{1/2}} = \frac{T^{1/2}\left(\widehat{\beta}_1 - 0\right)}{\left[\widehat{\sigma}^2 \frac{4T+2}{(T-1)}\right]^{1/2}} \\ &\Rightarrow \frac{d_1 - 6\sigma\left[\frac{1}{3}W(1) - \int_0^1 W(r)\,dr\right]}{\left[4\sigma^2\phi(b,k)\right]^{1/2}} = \frac{\dfrac{d_1}{2\sigma} - 3\left[\frac{1}{3}W(1) - \int_0^1 W(r)\,dr\right]}{\sqrt{\phi(b,k)}}, \end{aligned}$$

$$\begin{aligned} t_2 &= \frac{\left(\widehat{\beta}_2 - 0\right)}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{22}\right]^{1/2}} = \frac{T^{3/2}\left(\widehat{\beta}_2 - 0\right)}{\left[\widehat{\sigma}^2 \frac{12T^3}{T^3-T}\right]^{1/2}} \\ &\Rightarrow \frac{-d_1 + 12\sigma\left[\frac{1}{2}W(1) - \int_0^1 W(r)\,dr\right]}{\left[12\sigma^2\phi(b,k)\right]^{1/2}} = \frac{-\dfrac{d_1}{\sqrt{12}\sigma} + \sqrt{12}\left[\frac{1}{2}W(1) - \int_0^1 W(r)\,dr\right]}{\sqrt{\phi(b,k)}}, \end{aligned}$$

and, if $\{u_t\}$ is $I(1)$,

$$\begin{aligned} t_1 &= \frac{\left(\widehat{\beta}_1 - 0\right)}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{11}\right]^{1/2}} = \frac{T^{-1/2}\left(\widehat{\beta}_1 - 0\right)}{\left[T^{-2}\widehat{\sigma}^2 \frac{4T+2}{(T-1)}\right]^{1/2}} \\ &\Rightarrow \frac{d_1 + 6d(1)\left[\frac{1}{3}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]}{\left[4d(1)^2\phi(b,k)\right]^{1/2}} = \frac{\dfrac{d_1}{2d(1)} + 3\left[\frac{1}{3}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]}{\sqrt{\phi(b,k)}}, \end{aligned}$$

$$\begin{aligned} t_2 &= \frac{\left(\widehat{\beta}_2 - 0\right)}{\left[\widehat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}_{22}\right]^{1/2}} = \frac{T^{1/2}\left(\widehat{\beta}_2 - 0\right)}{\left[T^{-2}\widehat{\sigma}^2 \frac{12T^3}{T^3-T}\right]^{1/2}} \\ &\Rightarrow \frac{-d_1 - 12d(1)\left[\frac{1}{2}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]}{\left[12d(1)^2\phi(b,k)\right]^{1/2}} \\ &= \frac{-\dfrac{d_1}{\sqrt{12}d(1)} - \sqrt{12}\left[\frac{1}{2}\int_0^1 V_{\overline{\alpha}}(r)\,dr - \int_0^1 rV_{\overline{\alpha}}(r)\,dr\right]}{\sqrt{\phi(b,k)}}. \end{aligned}$$

The results in the theorem follow by definition of δ.
■

B. List of Kernels and Their Second Derivatives

The kernels used in the paper are:

$$\text{Bartlett} \qquad k(x) = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Parzen } (a) \qquad k(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{if } |x| \leq \frac{1}{2} \\ 2(1 - |x|)^3, & \text{if } \frac{1}{2} < |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Quadratic Spectral } (QS) \qquad k(x) = \frac{25}{12\pi^2 x^2}\left[\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right]$$

$$\text{Daniell} \qquad k(x) = \frac{\sin(\pi x)}{\pi x}$$

$$\text{Bohman} \qquad k(x) = \begin{cases} (1 - x)\cos(\pi x) + \dfrac{\sin(\pi x)}{\pi}, & \text{if } |x| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

The second derivatives of kernels used in the paper are:

$$\text{Parzen } (a) \qquad k''(x) = \begin{cases} -12 + 36|x|, & \text{if } |x| \leq \frac{1}{2} \\ 12(1 - |x|), & \text{if } \frac{1}{2} < |x| \leq 1 \end{cases}$$

$$\text{QS} \qquad k''(x) = \begin{cases} -36\pi^2/125 & \text{if } x = 0 \\ \frac{125}{72\pi^3 x^5}\left[\left(12 - \frac{36\pi^2 x^2}{5}\right)\sin\left(\frac{6\pi x}{5}\right) + \left(\frac{216\pi^3 x^3}{125} - \frac{72\pi x}{5}\right)\cos\left(\frac{6\pi x}{5}\right)\right], & \text{otherwise} \end{cases}$$

$$\text{Daniell} \qquad k''(x) = \begin{cases} -\pi^2/3, & \text{if } x = 0 \\ \left(2/\pi x^3\right)\left[\sin(\pi x) - \pi x\cos(\pi x)\right] - (\pi/x)\sin(\pi x), & \text{otherwise} \end{cases}$$

$$\text{Bohman} \qquad k''(x) = \pi\sin(\pi x) - \pi^2(1 - x)\cos(\pi x)$$

C. Tables and Figures

TABLE 1. Empirical Null Hypothesis Rejection Probabilities: *PLR vs.Ad Hoc* Procedure♦

	Nominal Size											
	.010		.025		.050		.100		.150		.200	
T	*Ad Hoc*	*PLR*	*Ad Hoc*	*PLR*	*Ad Hoc*	*PLR*	*Ad Hoc*	*PLR*	*Ad Hoc*	*PLR*	*Ad Hoc*	*PLR*
50	.0122	.0127	.0307	.0290	.0637	.0553	.1342	.1060	.2099	.1563	.2897	.2062
100	.0108	.0113	.0287	.0269	.0608	.0523	.1307	.1027	.2069	.1529	.2872	.2032
200	.0101	.0106	.0276	.0260	.0597	.0514	.1299	.1019	.2054	.1516	.2853	.2016
500	.0098	.0103	.0271	.0254	.0584	.0503	.1282	.1000	.2042	.1501	.2846	.2003
1000	.0097	.0101	.0266	.0249	.0581	.0500	.1285	.1005	.2047	.1504	.2849	.2009

♦ DGP: $y_t = \beta_1 + \beta_2 t + u_t$, $\{u_t\}$ are i.i.d. $\mathbb{N}(0,1)$; OLS was used to obtain estimates of β_1 and β_2. For PLR, rejection probabilities were computed by using the asymptotic critical values from Table 2 (σ^2 known). When computing rejection probabilities using the *ad hoc* procedure, there were two steps involved: At the first step, signs of t_1 and t_2 were checked. If they were found to be of opposite signs, then it was proceeded to step two, otherwise, the null hypothesis was not rejected. At step two, the standard normal critical values were used (at the nominal level indicated) to carry out each one sided t-test. The null hypothesis was rejected whenever both of the one-sided t-tests were found significant at the indicated nominal level; For both the PLR test and the *ad hoc* procedure, σ^2 was estimated by $s^2 = \frac{1}{T-2}\sum_{t=1}^{T} \widehat{u}_t^2$, where $\{\widehat{u}_t\}_{t=1}^{T}$ are the OLS residuals; 1,000,000 replications were performed for each sample size.

TABLE 2. Asymptotic Right-Tail Critical Values ♦

% Point	PLR σ^2 known	$PLR_{BG} - Daniell$ $b = 0.02$	$b = 0.12$	$b = 0.22$
70.0	0.652	.7127	1.002	1.459
		(394.7)	(98.96)	(45.76)
75.0	0.844	.8941	1.304	1.954
		(418.4)	(103.8)	(47.59)
80.0	1.096	1.154	1.700	2.633
		(434.9)	(109.0)	(46.58)
85.0	1.443	1.517	2.242	3.682
		(457.0)	(117.9)	(50.59)
90.0	1.964	2.081	3.234	5.732
		(495.6)	(125.6)	(47.11)
91.0	2.106	2.210	3.568	6.306
		(515.6)	(123.5)	(46.56)
92.0	2.265	2.391	3.815	6.958
		(527.2)	(125.3)	(48.29)
93.0	2.448	2.585	4.155	7.863
		(546.4)	(126.5)	(47.63)
94.0	2.664	2.829	4.667	8.988
		(565.5)	(128.4)	(47.71)
95.0	2.923	3.146	5.253	10.46
		(583.4)	(132.0)	(50.93)
96.0	3.247	3.501	5.925	12.51
		(614.5)	(141.4)	(51.60)
97.0	3.668	4.029	6.879	14.96
		(643.9)	(151.1)	(55.66)
97.5	3.940	4.361	7.615	16.99
		(644.8)	(158.1)	(56.04)
98.0	4.276	4.646	8.513	19.08
		(691.4)	(164.1)	(65.26)
99.0	5.337	5.749	11.83	26.51
		(769.1)	(173.4)	(121.5)
99.5	6.433	6.956	15.39	39.27
		(906.2)	(194.6)	(121.5)

♦ At each significance level, the scaling constant c_{BG} is reported in parenthesis under the corresponding critical value. Critical values were calculated by Monte Carlo simulation methods: $\mathbb{N}(0,1)$ i.i.d. random deviates were used to approximate the Wiener processes in the limiting distributions of PLR_{BG} tests. The integrals were approximated by the normalized sums of 1,000 steps using 10,000 replications. In the case of known σ^2, critical values for PLR were computed by using the bivariate normal relationship between t_1 and t_2, and performing 10,000,000 replications.

TABLE 3. Empirical Null Hypothesis Rejection Probabilities ♦
(10% Nominal Level, $T = 50$)

		PLR			$PLR_{BG} - Daniell$		
θ	ρ	HAC	$HACPW$	$PARM$	$b = 0.02$	$b = 0.12$	$b = 0.22$
-.8	-.5	.000	.005	.041	.000	.031	.047
	-.3	.001	.002	.032	.000	.032	.053
	.0	.002	.002	.028	.000	.032	.059
	.3	.004	.004	.025	.000	.035	.061
	.5	.011	.013	.027	.001	.039	.065
	.7	.054	.066	.055	.010	.059	.075
	.9	.254	.254	.258	.097	.144	.132
	.95	.306	.297	.306	.128	.171	.148
	.99	.282	.277	.284	.118	.158	.132
	1.00	.267	.261	.268	.113	.152	.128
-.4	-.5	.066	.075	.104	.000	.091	.093
	-.3	.068	.068	.104	.001	.090	.094
	.0	.072	.077	.096	.004	.087	.095
	.3	.101	.115	.097	.023	.086	.094
	.5	.163	.158	.169	.054	.086	.095
	.7	.253	.223	.242	.089	.100	.100
	.9	.353	.324	.302	.107	.134	.123
	.95	.355	.328	.298	.093	.131	.120
	.99	.316	.295	.266	.077	.117	.108
	1.00	.309	.289	.260	.077	.118	.109
.0	-.5	.114	.135	.133	.005	.096	.098
	-.3	.111	.131	.117	.014	.092	.097
	.0	.126	.134	.126	.036	.086	.094
	.3	.176	.142	.174	.064	.080	.091
	.5	.205	.153	.168	.076	.078	.090
	.7	.256	.169	.182	.086	.085	.091
	.9	.339	.213	.220	.073	.104	.105
	.95	.342	.215	.223	.059	.099	.101
	.99	.315	.198	.204	.052	.090	.095
	1.00	.313	.196	.204	.051	.092	.094
.4	−.5	.124	.142	.118	.030	.088	.095
	−.3	.133	.128	.139	.042	.084	.094
	.0	.155	.110	.158	.060	.078	.091
	.3	.187	.096	.162	.069	.072	.088
	.5	.217	.092	.179	.075	.070	.087
	.7	.269	.101	.209	.076	.075	.088
	.9	.367	.130	.256	.060	.091	.099
	.95	.373	.133	.256	.050	.088	.096
	.99	.341	.126	.231	.044	.080	.088
	1.00	.342	.127	.234	.045	.083	.089
.8	-.5	.134	.108	.158	.047	.082	.092
	-.3	.142	.094	.174	.055	.079	.091
	.0	.164	.077	.192	.062	.074	.089
	.3	.193	.071	.211	.068	.069	.087
	.5	.225	.070	.227	.072	.067	.085
	.7	.278	.078	.251	.072	.072	.086
	.9	.378	.108	.298	.057	.086	.095
	.95	.386	.112	.295	.048	.082	.092
	.99	.354	.107	.267	.043	.078	.085
	1.00	.356	.109	.266	.044	.081	.087

♦ DGP: $y_t = \beta_1 + \beta_2 t + u_t$, $u_t = \rho u_{t-1} + e_t + \theta e_{t-1}$, $\{e_t\}$ are i.i.d. $\mathbb{N}(0,1)$, $u_0 = 0$, $e_0 = 0$; Rejection probabilities were computed by using asymptotic 10% critical values (from Table 2); 10,000 replications were performed for each ARMA$(1,1)$ specification.

TABLE 4. Empirical Null Hypothesis Rejection Probabilities ♦
(10% Nominal Level, $T = 100$)

θ	ρ	PLR			$PLR_{BG} - Daniell$		
		HAC	$HACPW$	$PARM$	$b = 0.02$	$b = 0.12$	$b = 0.22$
-.8	-.5	.000	.003	.026	.000	.077	.074
	-.3	.000	.001	.019	.000	.076	.079
	.0	.001	.001	.017	.000	.073	.082
	.3	.003	.001	.015	.000	.071	.083
	.5	.007	.005	.016	.001	.070	.085
	.7	.038	.045	.035	.019	.079	.090
	.9	.279	.272	.285	.157	.140	.119
	.95	.386	.379	.379	.209	.182	.140
	.99	.366	.364	.347	.182	.177	.137
	1.00	.337	.333	.319	.167	.163	.126
-.4	-.5	.059	.062	.087	.020	.102	.099
	-.3	.063	.057	.084	.021	.102	.098
	.0	.070	.061	.082	.028	.100	.097
	.3	.089	.098	.082	.051	.097	.098
	.5	.148	.139	.155	.080	.097	.097
	.7	.222	.207	.204	.116	.095	.097
	.9	.327	.333	.263	.124	.111	.104
	.95	.356	.367	.279	.108	.116	.109
	.99	.316	.319	.244	.077	.104	.098
	1.00	.302	.307	.236	.075	.102	.098
.0	-.5	.103	.116	.117	.063	.102	.098
	-.3	.100	.115	.109	.061	.102	.098
	.0	.112	.116	.109	.065	.097	.097
	.3	.150	.121	.138	.076	.092	.095
	.5	.174	.125	.129	.086	.086	.094
	.7	.207	.139	.143	.096	.082	.092
	.9	.296	.181	.187	.088	.092	.095
	.95	.337	.202	.208	.073	.097	.097
	.99	.310	.187	.191	.054	.087	.088
	1.00	.301	.180	.185	.053	.086	.088
.4	−.5	.108	.125	.103	.067	.099	.098
	−.3	.118	.110	.126	.063	.096	.097
	.0	.136	.090	.122	.062	.092	.095
	.3	.158	.074	.135	.070	.085	.093
	.5	.179	.068	.153	.077	.080	.093
	.7	.215	.069	.171	.085	.076	.089
	.9	.311	.089	.219	.080	.086	.092
	.95	.364	.111	.237	.065	.090	.093
	.99	.338	.113	.217	.048	.082	.088
	1.00	.333	.109	.208	.047	.080	.086
.8	-.5	.119	.091	.135	.058	.095	.096
	-.3	.130	.077	.146	.055	.093	.096
	.0	.144	.065	.159	.057	.089	.094
	.3	.163	.059	.170	.066	.083	.094
	.5	.185	.054	.176	.074	.078	.092
	.7	.220	.052	.192	.081	.075	.089
	.9	.324	.070	.239	.078	.084	.091
	.95	.376	.089	.255	.063	.089	.092
	.99	.351	.092	.233	.047	.080	.086
	1.00	.344	.090	.223	.046	.079	.085

♦ DGP: $y_t = \beta_1 + \beta_2 t + u_t$, $u_t = \rho u_{t-1} + e_t + \theta e_{t-1}$, $\{e_t\}$ are i.i.d. $\mathbb{N}(0,1)$, $u_0 = 0$, $e_0 = 0$; Rejection probabilities were computed by using asymptotic 10% critical values (from Table 2); 10,000 replications were performed for each ARMA$(1,1)$ specification.

TABLE 5. Empirical Null Hypothesis Rejection Probabilities ♦
(10% Nominal Level, $T = 200$)

		PLR			$PLR_{BG} - Daniell$		
θ	ρ	*HAC*	*HACPW*	*PARM*	$b = 0.02$	$b = 0.12$	$b = 0.22$
-.8	-.5	.000	.002	.013	.003	.095	.084
	-.3	.001	.001	.011	.003	.094	.089
	.0	.001	.000	.011	.002	.093	.092
	.3	.003	.001	.014	.004	.091	.093
	.5	.006	.003	.016	.009	.089	.094
	.7	.031	.035	.029	.033	.089	.095
	.9	.287	.280	.290	.178	.116	.103
	.95	.416	.425	.371	.246	.150	.117
	.99	.449	.472	.378	.219	.176	.132
	1.00	.384	.404	.329	.183	.163	.122
-.4	-.5	.056	.058	.078	.068	.098	.097
	-.3	.059	.053	.075	.067	.099	.096
	.0	.066	.055	.072	.069	.097	.097
	.3	.079	.087	.073	.075	.096	.096
	.5	.134	.126	.143	.084	.093	.094
	.7	.186	.191	.167	.106	.089	.094
	.9	.269	.316	.201	.127	.089	.093
	.95	.325	.374	.230	.112	.098	.099
	.99	.343	.378	.240	.079	.104	.101
	1.00	.314	.339	.222	.064	.094	.092
.0	-.5	.093	.106	.105	.086	.097	.096
	-.3	.092	.105	.105	.083	.097	.096
	.0	.102	.106	.101	.078	.095	.095
	.3	.132	.108	.109	.075	.091	.094
	.5	.147	.110	.111	.078	.087	.093
	.7	.167	.114	.117	.088	.081	.091
	.9	.237	.145	.147	.094	.077	.089
	.95	.296	.174	.178	.083	.086	.093
	.99	.331	.195	.198	.061	.093	.094
	1.00	.307	.184	.188	.050	.084	.086
.4	$-.5$	.098	.113	.095	.080	.095	.096
	$-.3$	.108	.099	.114	.077	.094	.095
	.0	.121	.076	.103	.070	.092	.094
	.3	.136	.061	.121	.068	.088	.093
	.5	.147	.054	.131	.070	.083	.092
	.7	.171	.047	.140	.080	.078	.090
	.9	.249	.052	.171	.087	.074	.086
	.95	.317	.073	.202	.078	.082	.092
	.99	.368	.104	.213	.057	.089	.092
	1.00	.339	.105	.203	.046	.081	.085
.8	-.5	.111	.079	.117	.074	.093	.095
	-.3	.116	.063	.126	.070	.093	.095
	.0	.125	.059	.133	.067	.090	.093
	.3	.137	.052	.140	.065	.087	.092
	.5	.150	.044	.146	.068	.083	.092
	.7	.173	.035	.156	.076	.078	.089
	.9	.258	.037	.188	.086	.073	.086
	.95	.329	.053	.221	.077	.081	.091
	.99	.381	.086	.231	.055	.089	.091
	1.00	.352	.086	.220	.045	.081	.085

♦ DGP: $y_t = \beta_1 + \beta_2 t + u_t$, $u_t = \rho u_{t-1} + e_t + \theta e_{t-1}$, $\{e_t\}$ are i.i.d. $\mathbb{N}(0,1)$, $u_0 = 0$, $e_0 = 0$; Rejection probabilities were computed by using asymptotic 10% critical values (from Table 2); 10,000 replications were performed for each ARMA$(1,1)$ specification.

TABLE 6. The Distribution of US States in Eight Regions, as defined by the Bureau of Economic Analysis

New England:	**Mideast:**	**Great Lakes:**	**Plains:**	**Southeast:**	**Southwest:**	**Rocky Mountains:**	**Farwest:**
Connecticut	Delaware	Illinois	Iowa	Alabama	Arizona	Colorado	Alaska
Maine	Dist. of Columbia	Indiana	Kansas	Arkansas	New Mexico	Idaho	California
Massachusetts	Maryland	Michigan	Minnesota	Florida	Oklahoma	Montana	Hawaii
New Hampshire	New Jersey	Ohio	Missouri	Georgia	Texas	Utah	Nevada
Rhode Island	New York	Wisconsin	Nebraska	Kentucky		Wyoming	Oregon
Vermont	Pennsylvania		North Dakota	Louisiana			Washington
			South Dakota	Mississippi			
				North Carolina			
				South Carolina			
				Tennessee			
				Virginia			
				West Virgina			

TABLE 7. Empirical tests of US regional convergence using the model, $y_t = \beta_1 + \beta_2 t + u_t$

	New England		Mideast		Great Lakes		Plains		Southeast		Southwest		Rocky Mts		Farwest	
	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
OLS Estimates:	0.0442	0.00193	0.129	-0.000304	0.112	-0.00261	-0.0603	0.000252	-0.364	0.00547	-0.147	0.000972	-0.0270	-0.00121	0.212	-0.00305
Tests of H_0: $\beta_i = 0$																
2% t_J - Dan (b = 0.02)	0.000	0.000	0.000	-0.000	**4.631**	-0.621	**-2.719**	0.080	-0.000	0.000	-0.048	0.000	-0.122	-0.004	**2.719**	-0.049
2% t_{BG} - Dan (b = 0.16)	0.701	0.035	3.696	-0.015	**7.698**	-1.358	**-4.318**	0.193	**-5.090**	0.016	-3.069	0.044	-1.293	-0.601	**10.079**	-0.263
5% t_J - Dan (b = 0.02)	0.000	0.000	0.002	-0.000	**6.670**	-1.528	**-3.747**	0.177	-0.000	0.000	-0.185	0.000	-0.272	-0.028	**5.379**	-0.266
5% t_{BG} - Dan (b = 0.16)	0.756	0.097	**3.952**	-0.036	**7.957**	-2.116	**-4.428**	0.271	**-5.707**	0.073	**-3.262**	0.100	-1.327	-0.851	**10.759**	-0.630
10% t_J - Dan (b = 0.02)	0.000	0.000	0.009	-0.000	**8.305**	**-2.881**	**-4.543**	0.310	-0.000	0.000	-0.415	0.005	-0.440	-0.113	**8.106**	-0.870
10% t_{BG} - Dan (b = 0.16)	0.788	0.191	**4.101**	-0.065	**8.104**	**-2.841**	**-4.490**	0.339	**-6.080**	0.201	**-3.374**	0.173	-1.346	-1.072	**11.155**	-1.126
Joint Tests:																
2% PLR_{BG} - Dan (b = 0.02)	0.000		0.001		**5.050**		0.100		0.000		0.007		0.000		0.197	
2% PLR_{BG} - Dan (b = 0.12)	0.000		0.070		**33.29**		0.340		3.937		0.395		0.000		**20.32**	
2% PLR_{BG} - Dan (b = 0.22)	0.000		0.162		**55.36**		0.387		**20.06**		0.789		0.000		**58.27**	
5% PLR_{BG} - Dan (b = 0.02)	0.000		0.003		**9.497**		0.162		0.004		0.021		0.000		0.686	
5% PLR_{BG} - Dan (b = 0.12)	0.000		0.102		**40.17**		0.392		**7.532**		0.558		0.000		**29.43**	
5% PLR_{BG} - Dan (b = 0.22)	0.000		0.192		**60.20**		0.412		**26.80**		0.920		0.000		**68.75**	
10% PLR_{BG} - Dan (b = 0.02)	0.000		0.008		**15.87**		0.239		0.024		0.054		0.000		1.889	
10% PLR_{BG} - Dan (b = 0.12)	0.000		0.110		**41.70**		0.404		**8.572**		0.598		0.000		**31.68**	
10% PLR_{BG} - Dan (b = 0.22)	0.000		0.200		**61.56**		0.419		**28.95**		0.959		0.000		**71.84**	
Unit Root Tests:																
J	16.592		10.635		1.341##		1.179##		19.283		4.954		2.951		2.508	
BG	0.01332		0.01165		0.00585		0.00445#		0.02021		0.01078		0.00458#		0.01153	
β - Convergence:	**No Convergence**		**No Convergence**		**Convergence Occuring**		**No Convergence**		**Convergence Occuring**		**No Convergence**		**No Convergence**		**Convergence Occuring**	

Notes: Bold entries indicate significance at the level given in the left-most column. For unit root tests *J* and *BG*, "##", and "#" are used to indicate significance at 10% and 15% levels respectively. Critical values used in carrying out the joint tests can be found in Table 2. t_J and t_{BG} test statistics are reported to 3 decimal places. t_J and t_{BG} are unit root robust tests and they are implemented using the Daniell kernel and the bandwidth values as recommended by Bunzel and Vogelsang (2003). The critical values necessary for carrying out these tests were also obtained from that paper.

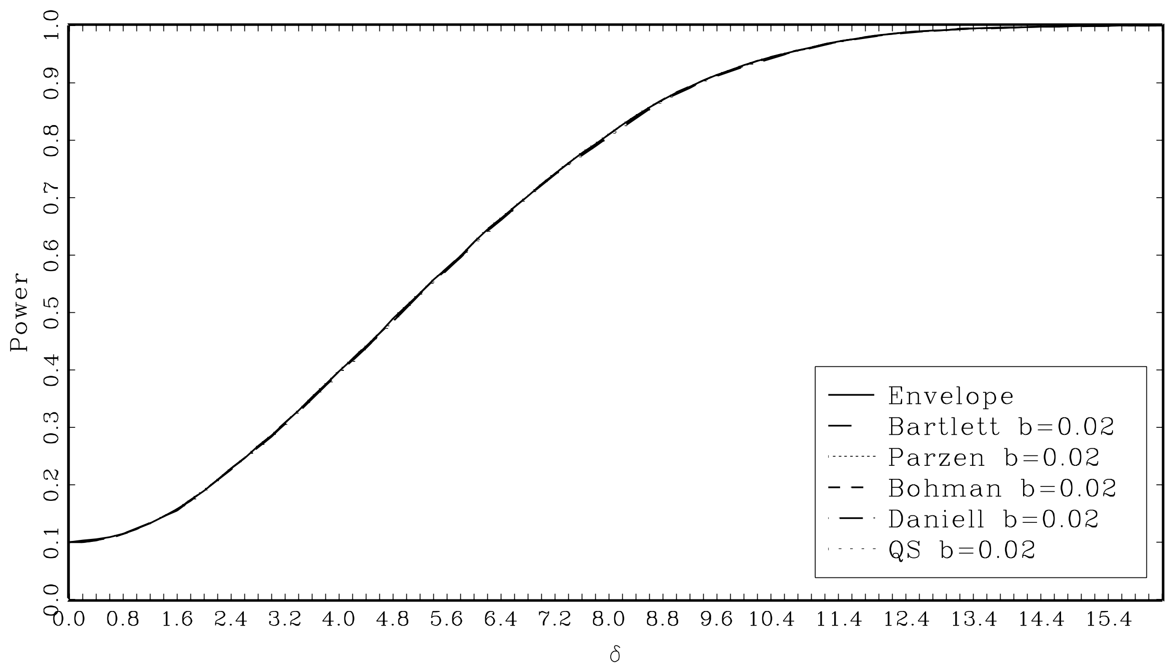


Figure 2: Asymptotic Power, 10% Asy. CVs Used, Stationary Errors

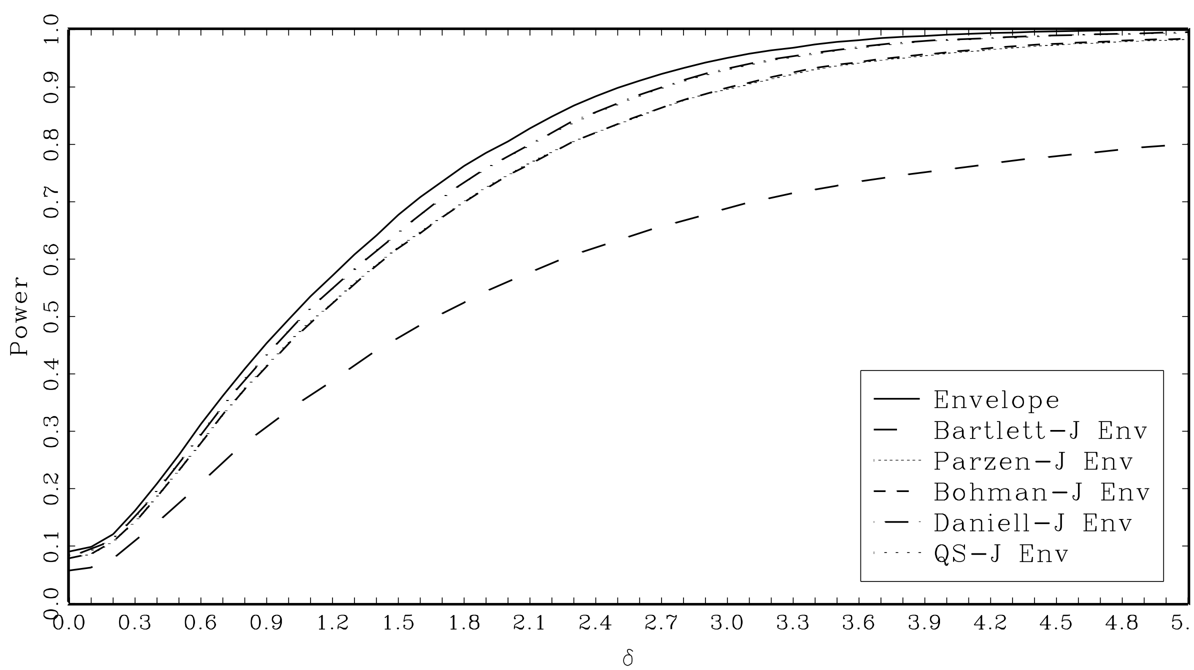


Figure 3: Asymptotic Power, 10% Asy. CVs Used, I(1) Errors, αbar=0

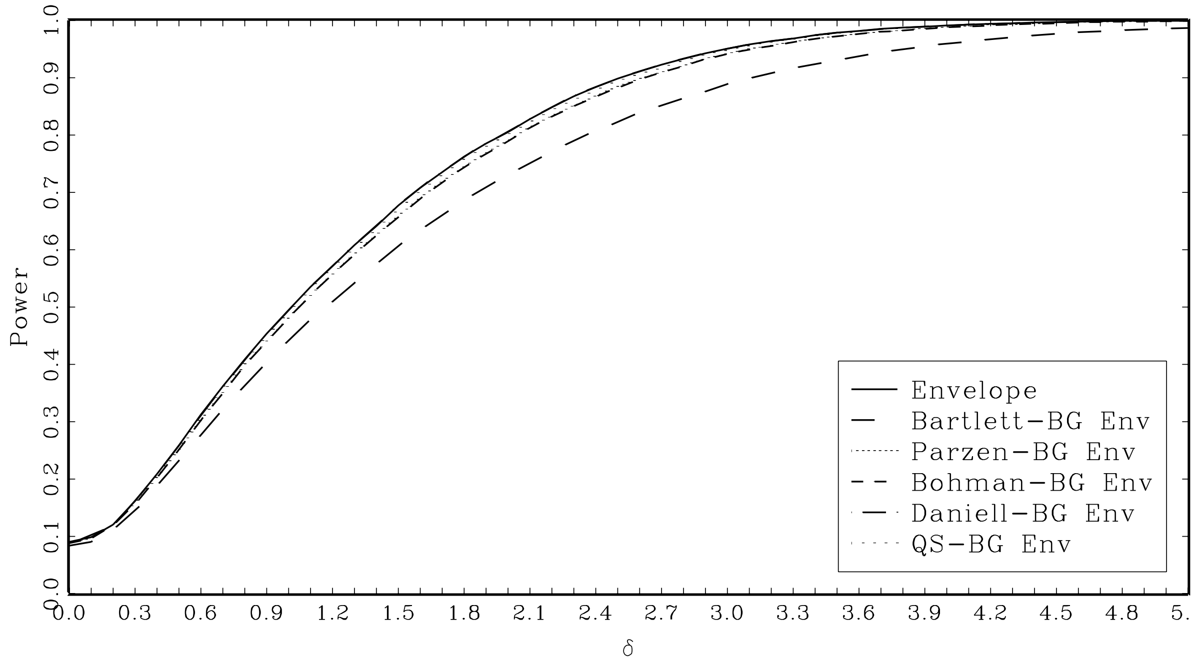


Figure 4: Asymptotic Power, 10% Asy. CVs Used, I(1) Errors, αbar=0

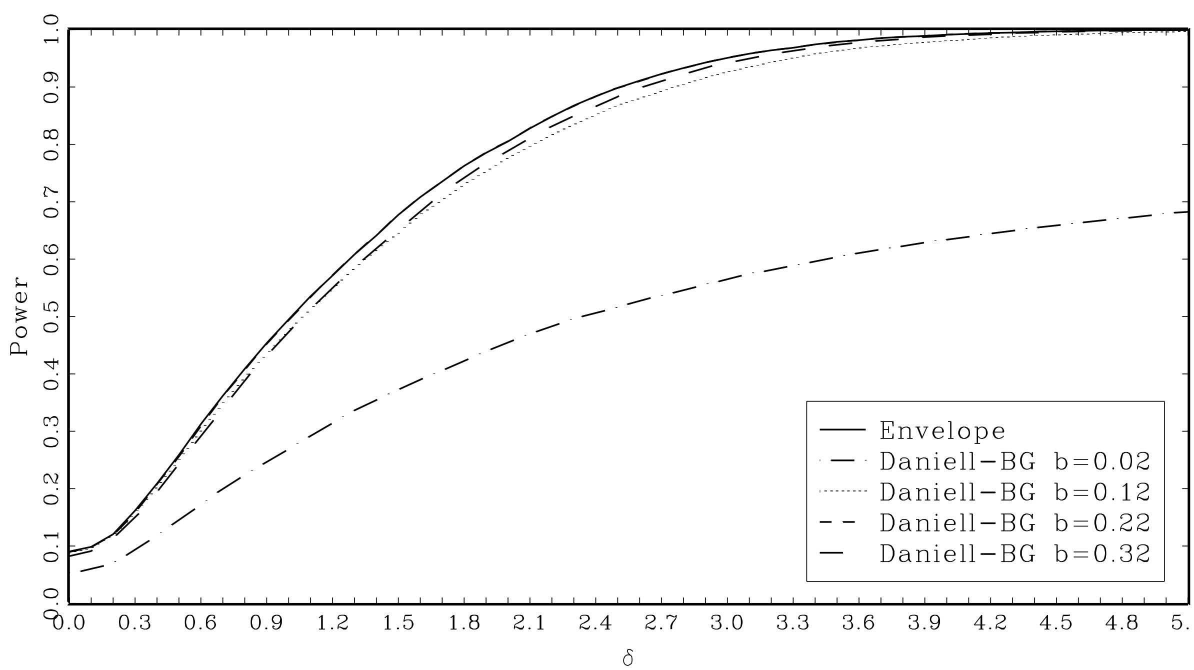


Figure 5: Asymptotic Power, 10% Asy. CVs Used, I(1) Errors, αbar=0

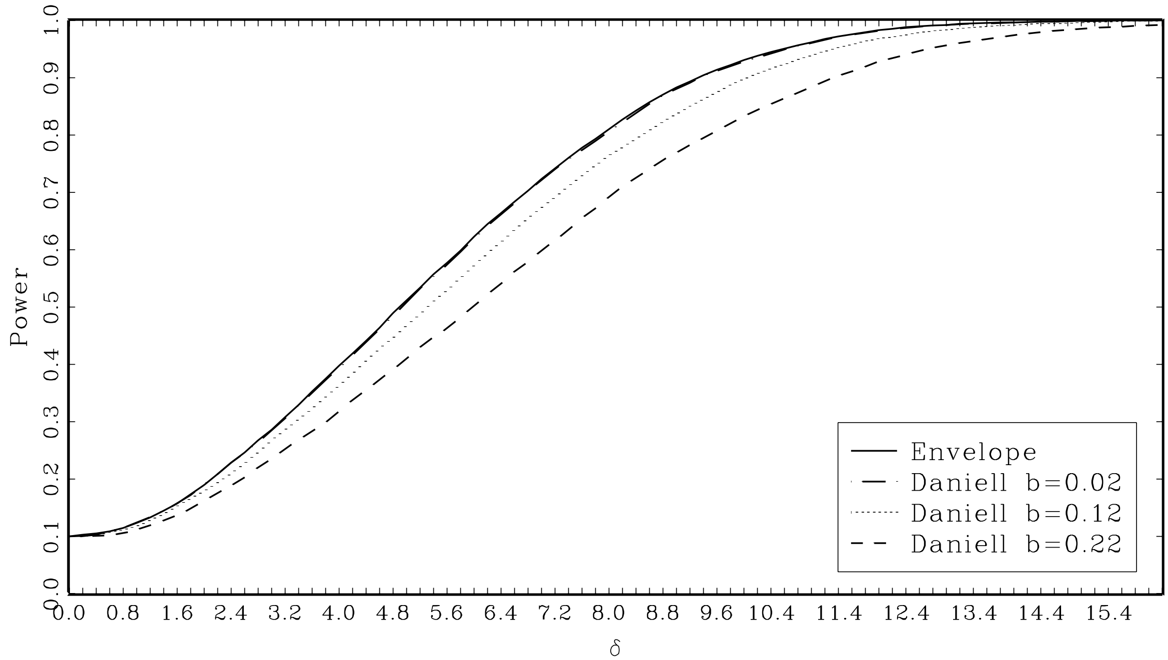


Figure 6: Asymptotic Power, 10% Asy. CVs Used, Stationary Errors

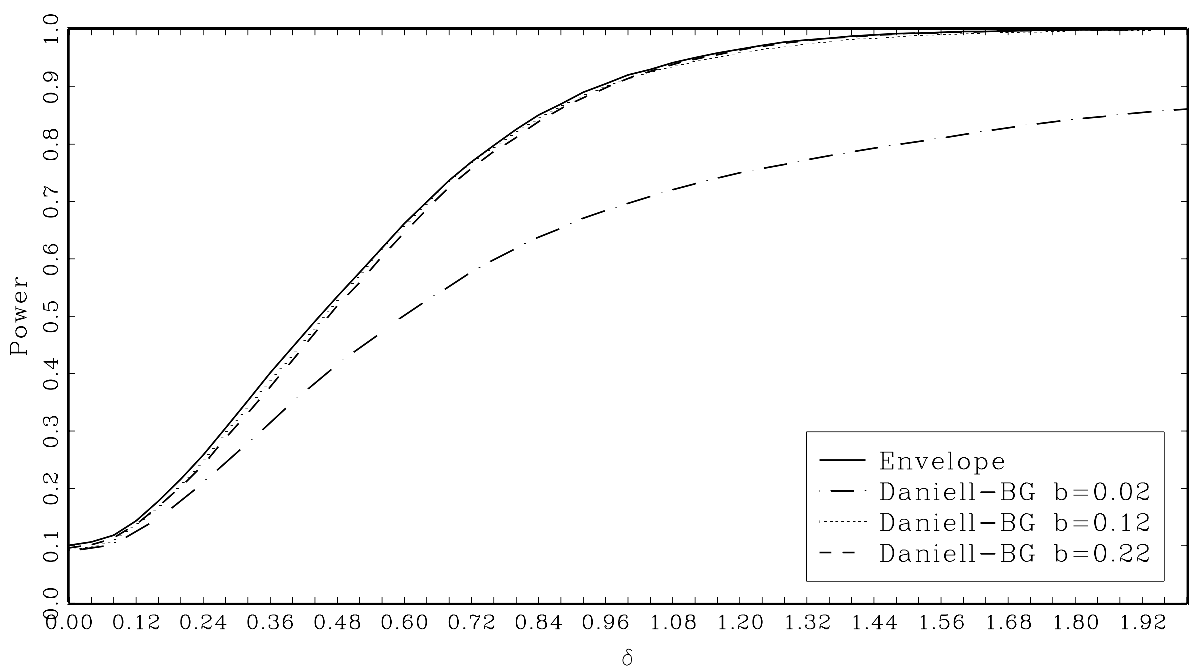


Figure 7: Asymptotic Power, 10% Asy. CVs Used, I(1) Errors, αbar=10

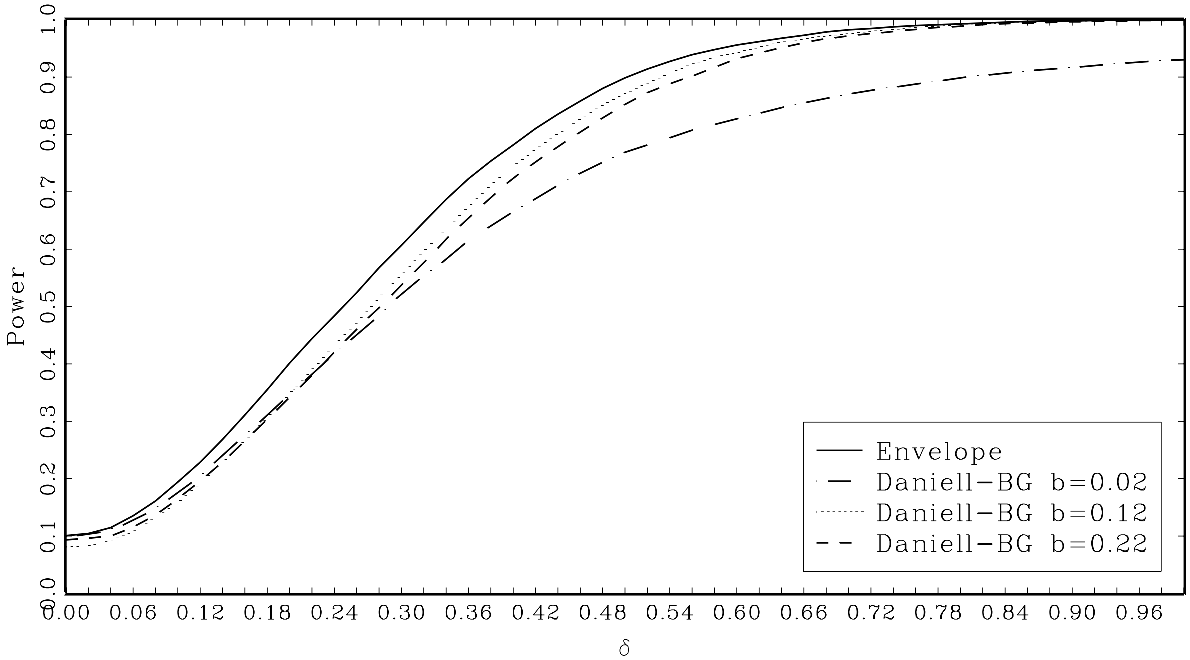


Figure 8: Asymptotic Power, 10% Asy. CVs Used, I(1) Errors, αbar=20

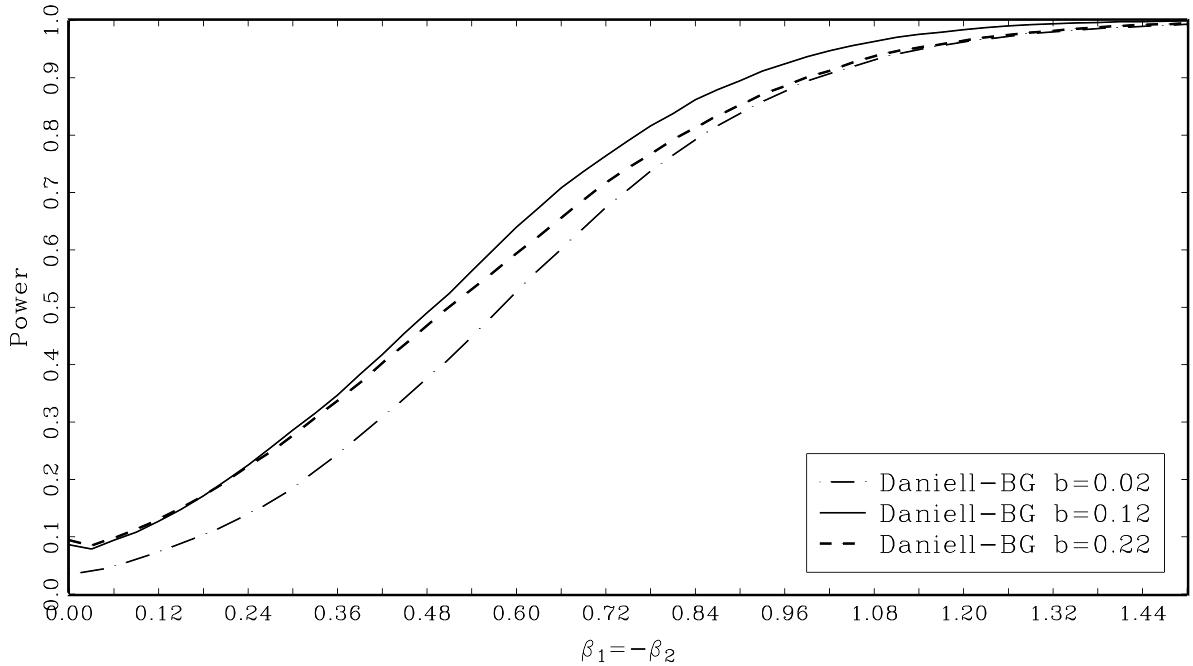


Figure 9: Finite Sample Power, AR(1) Errors, α=0.0, T=50

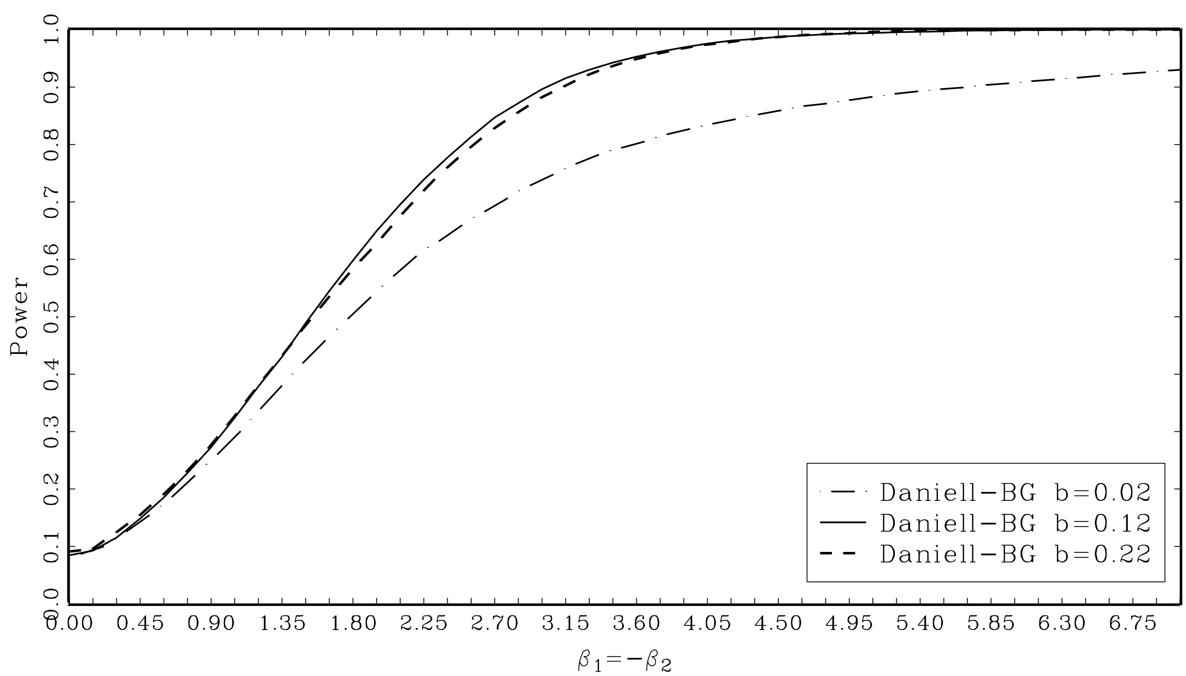


Figure 10: Finite Sample Power, AR(1) Errors, α=0.7, T=50

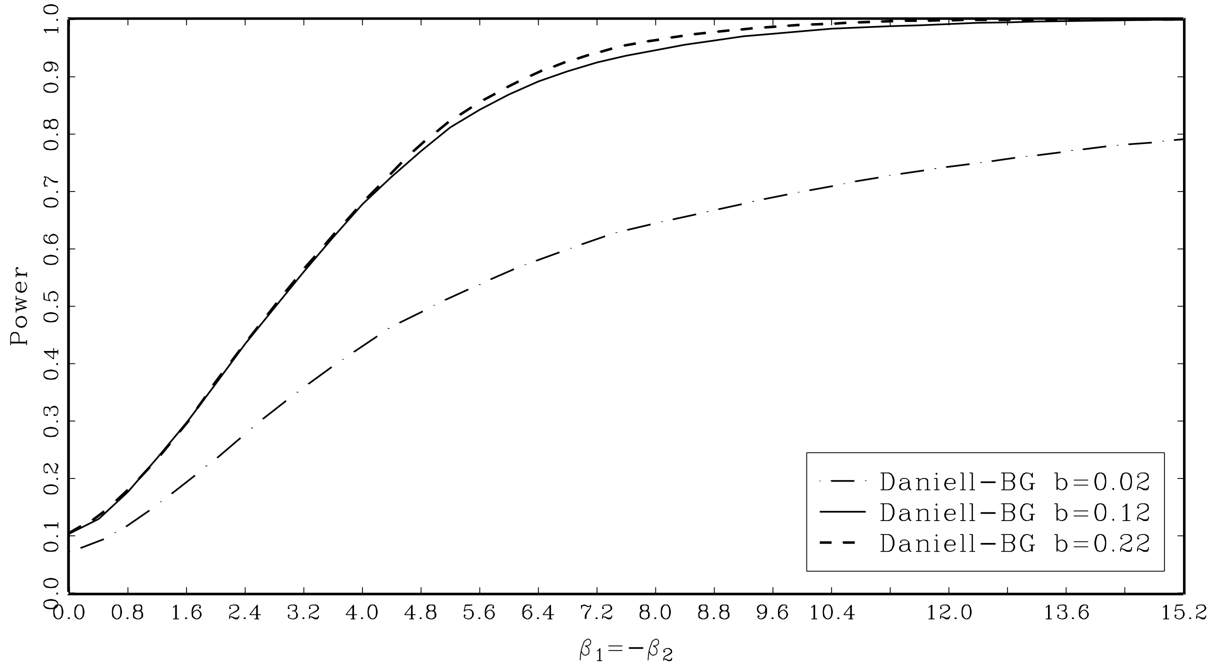


Figure 11: Finite Sample Power, AR(1) Errors, α=0.9, T=50

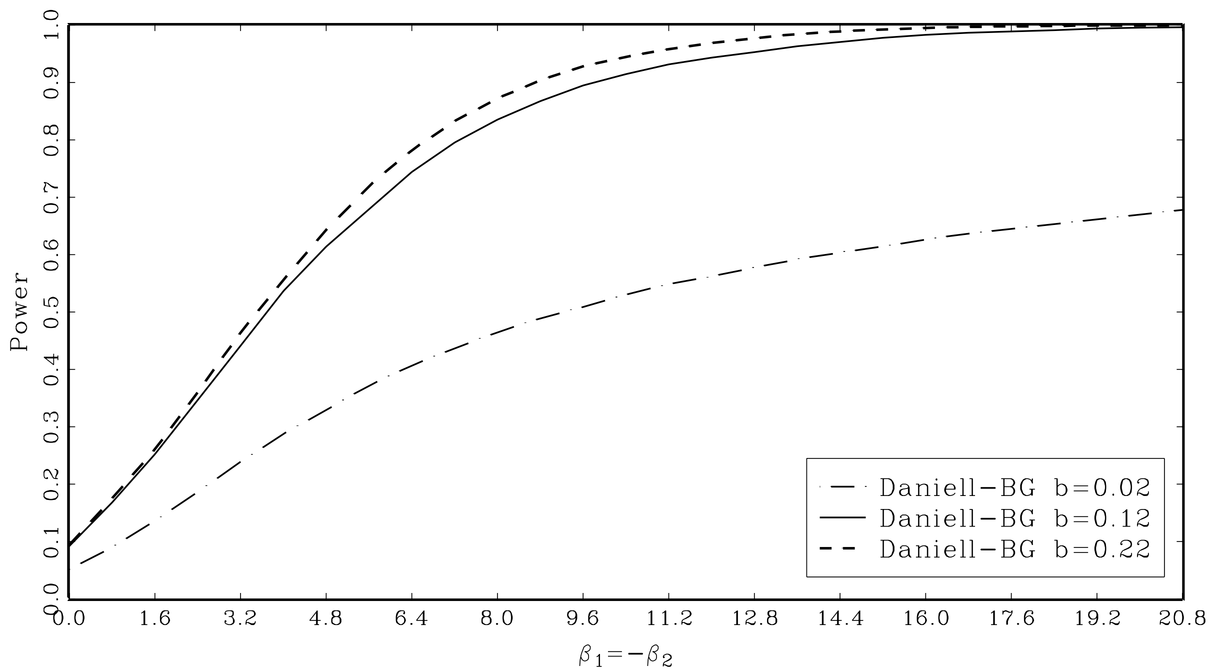


Figure 12: Finite Sample Power, AR(1) Errors, α=1.0, T=50

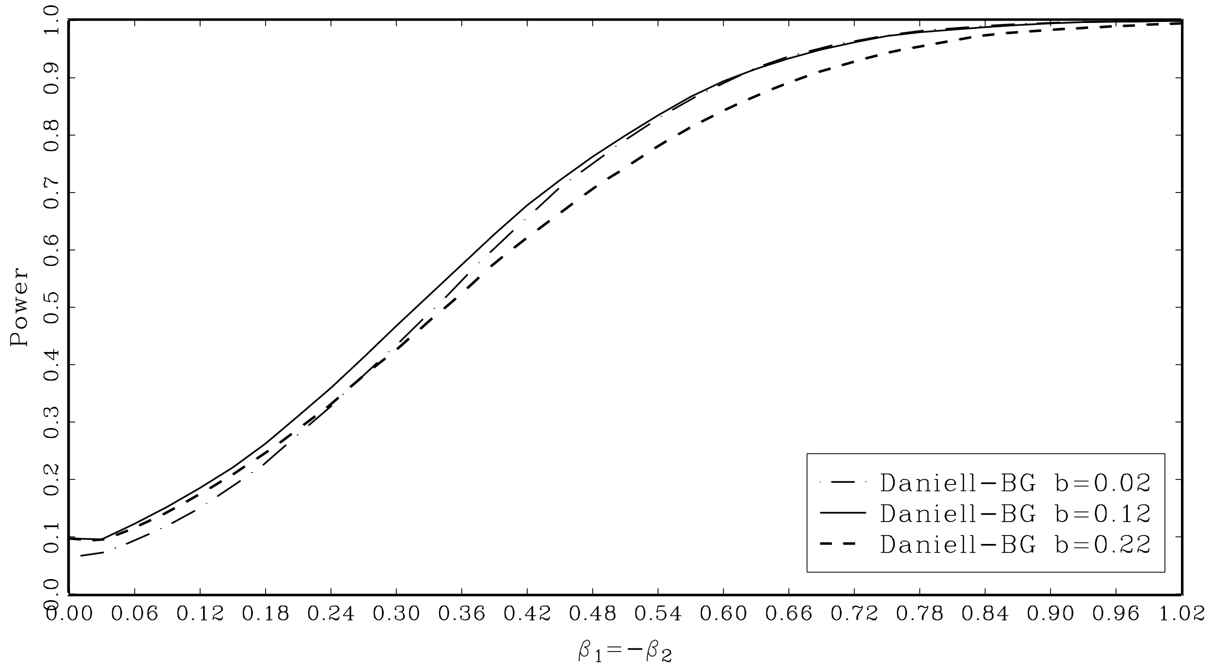


Figure 13: Finite Sample Power, AR(1) Errors, α=0.0, T=100

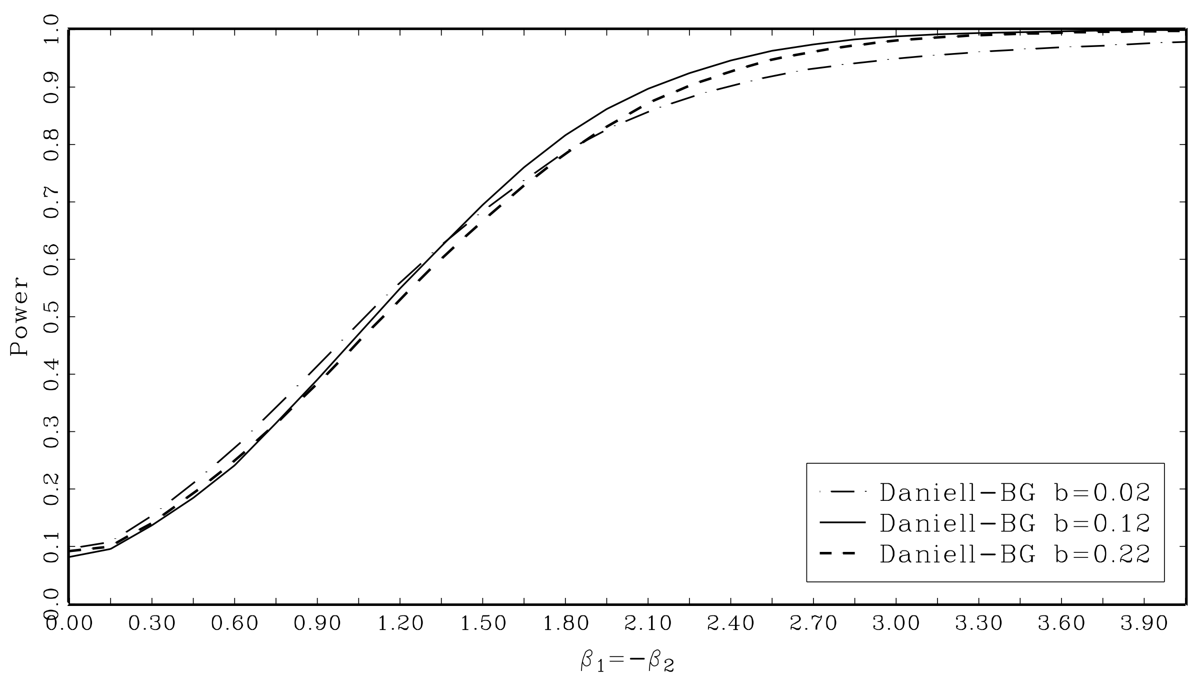


Figure 14: Finite Sample Power, AR(1) Errors, α=0.7, T=100

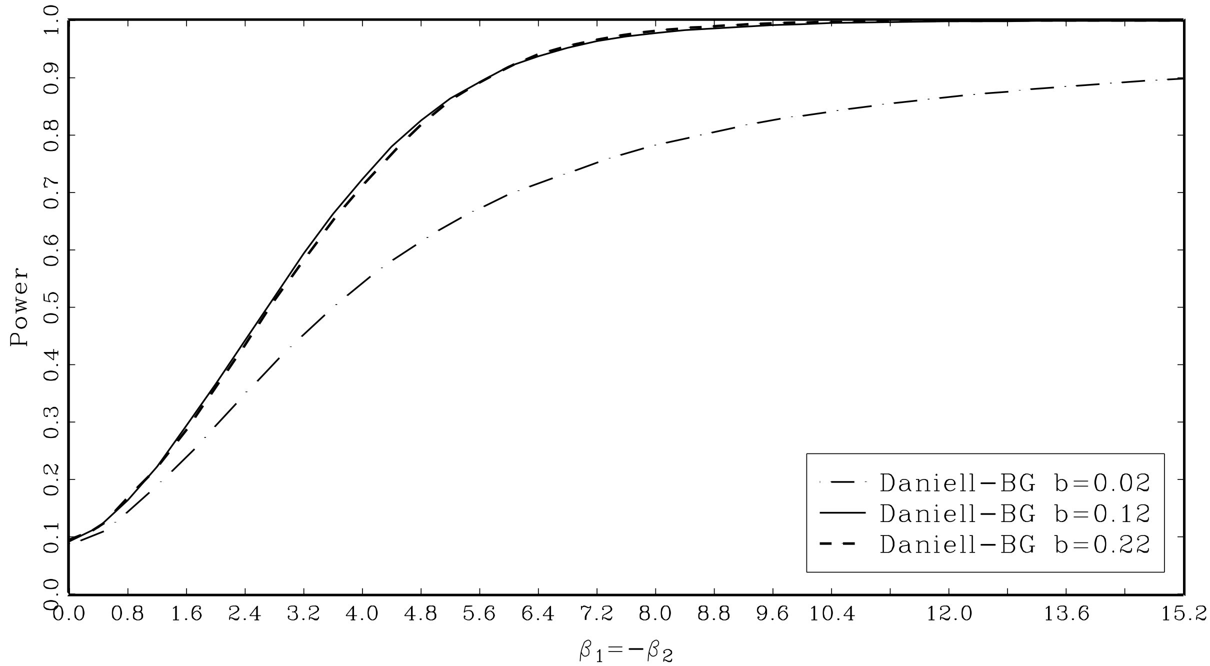


Figure 15: Finite Sample Power, AR(1) Errors, α=0.9, T=100

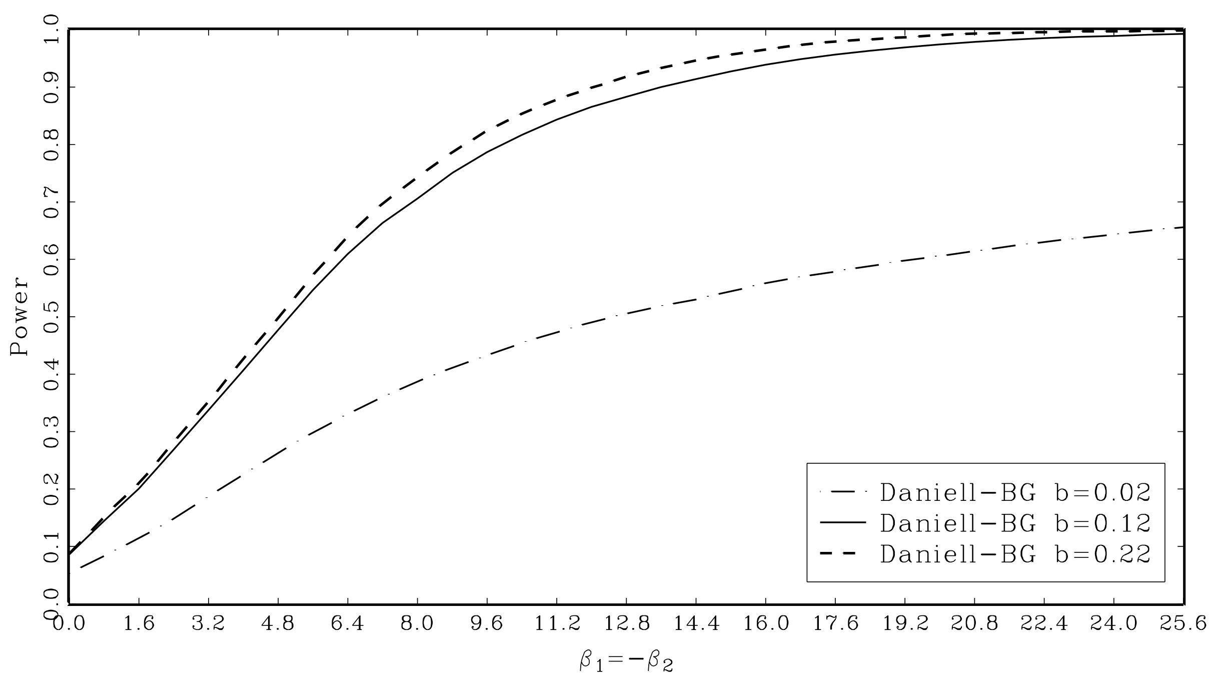


Figure 16: Finite Sample Power, AR(1) Errors, α=1.0, T=100

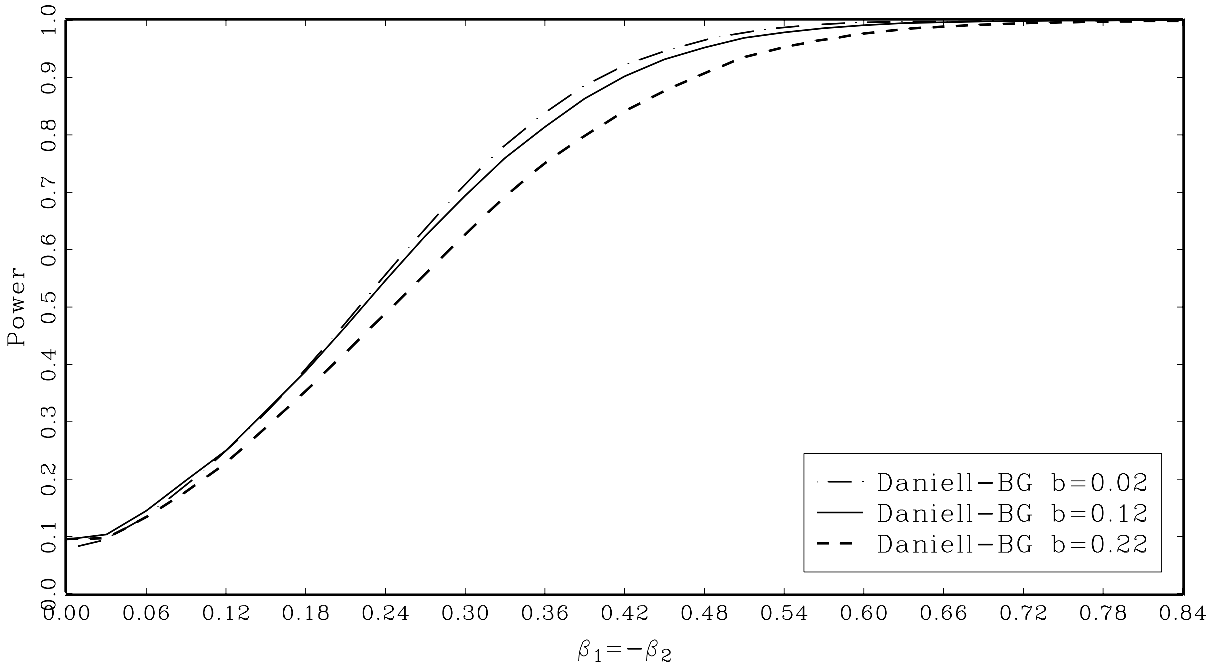


Figure 17: Finite Sample Power, AR(1) Errors, α=0.0, T=200

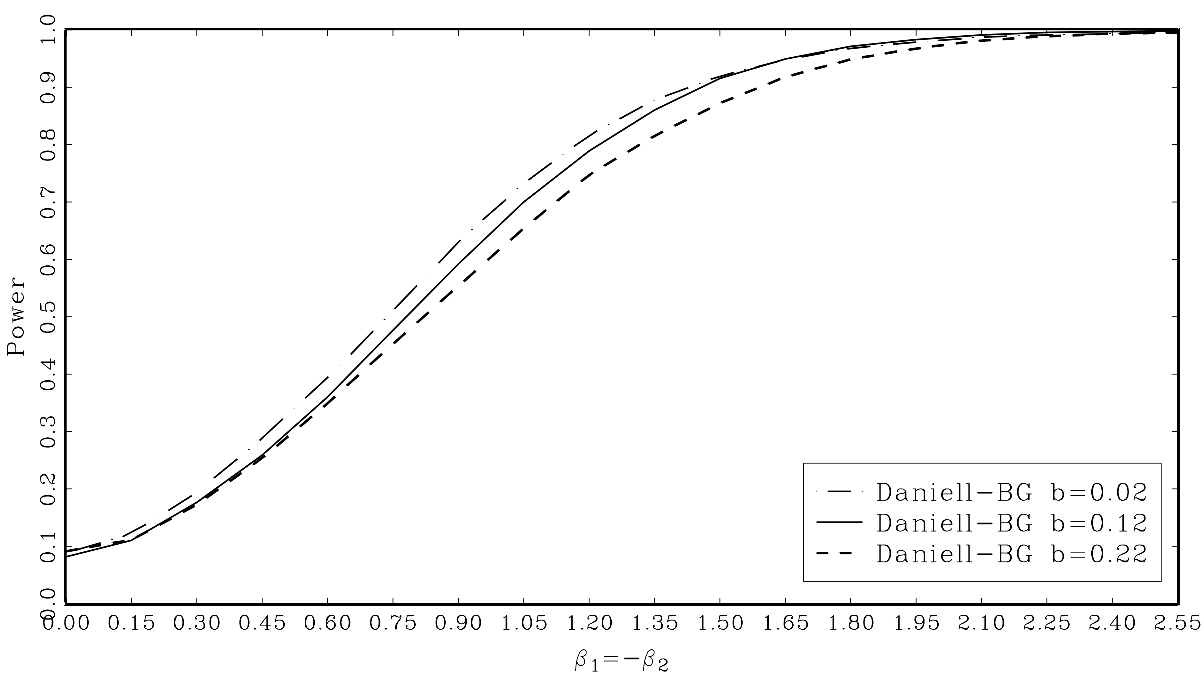


Figure 18: Finite Sample Power, AR(1) Errors, α=0.7, T=200

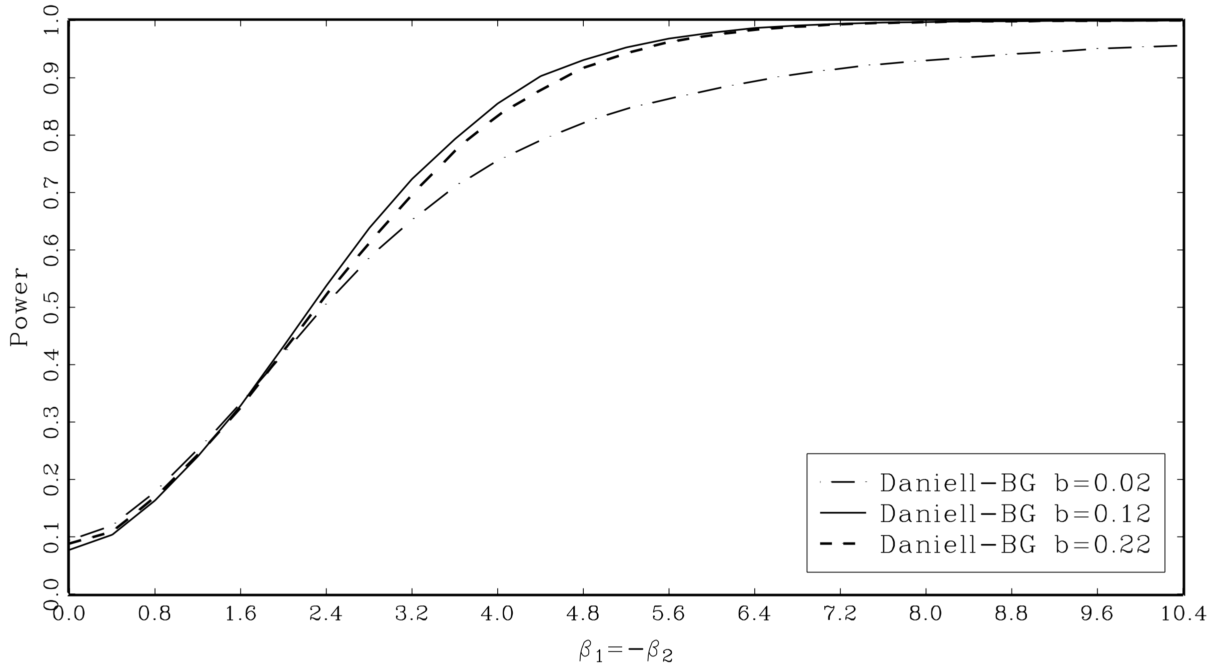


Figure 19: Finite Sample Power, AR(1) Errors, α=0.9, T=200

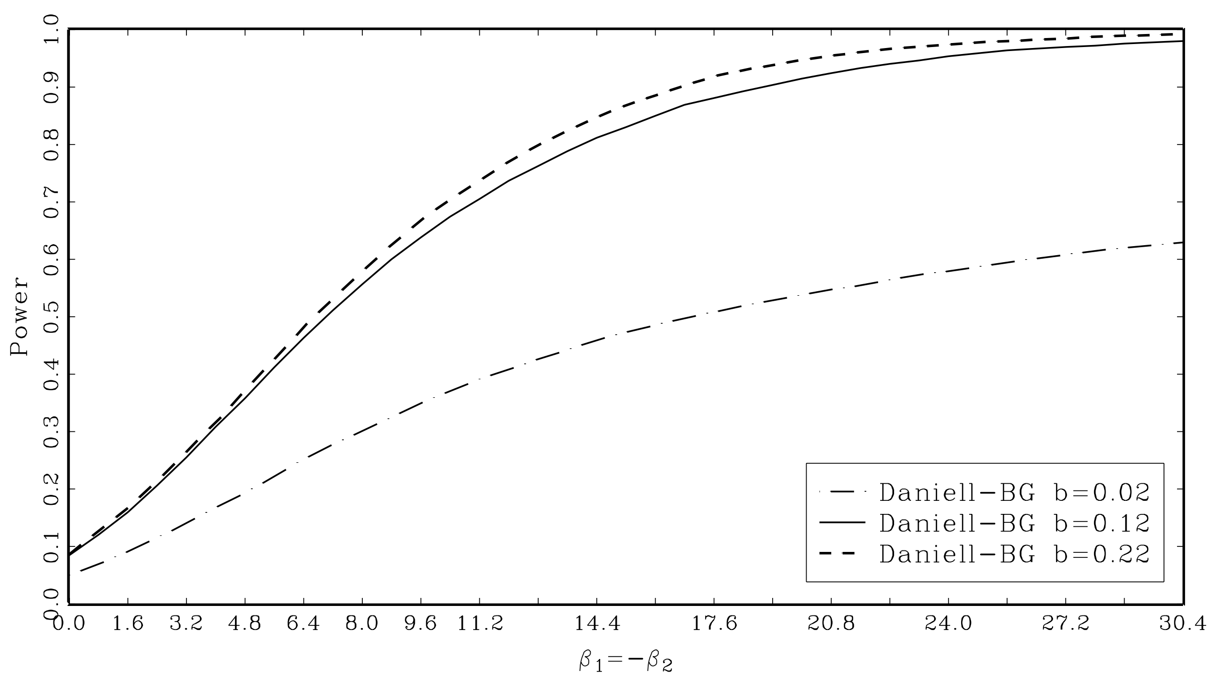


Figure 20: Finite Sample Power, AR(1) Errors, α=1.0, T=200

Figure 21. New England

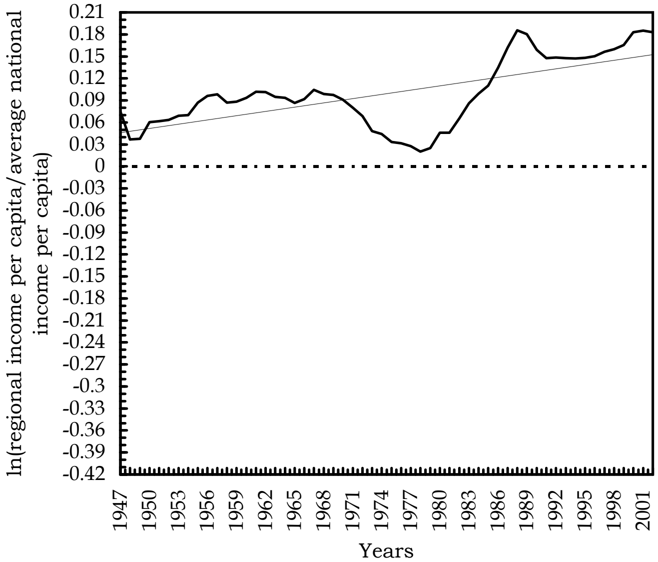


Figure 22. Mideast

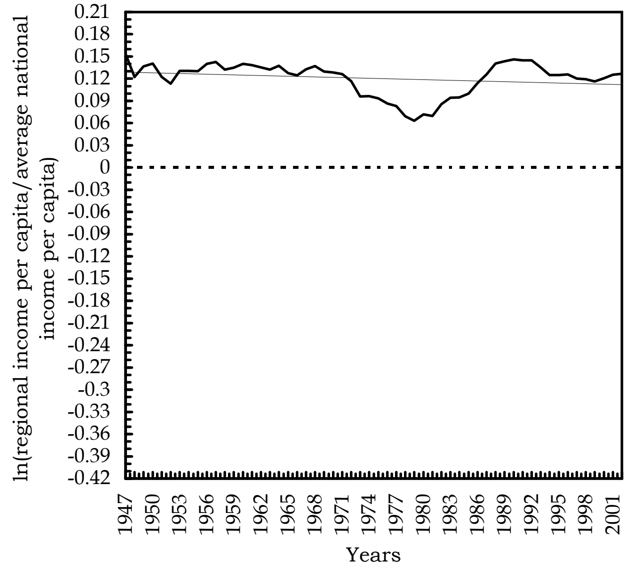


Figure 23. Great Lakes

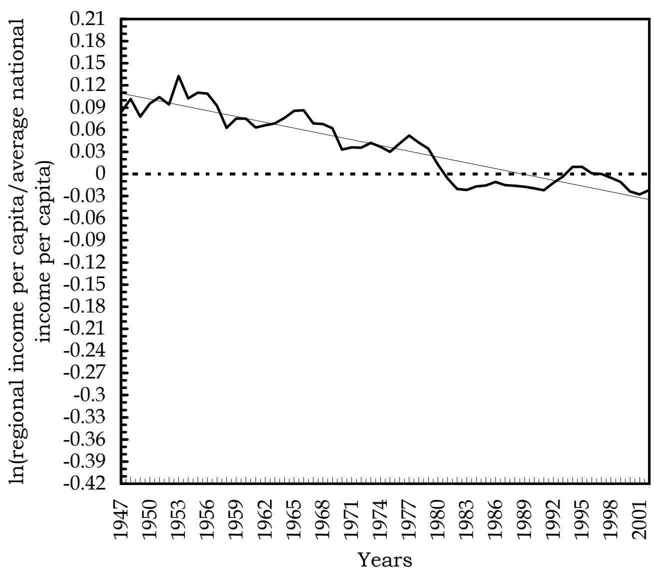


Figure 24. Plains

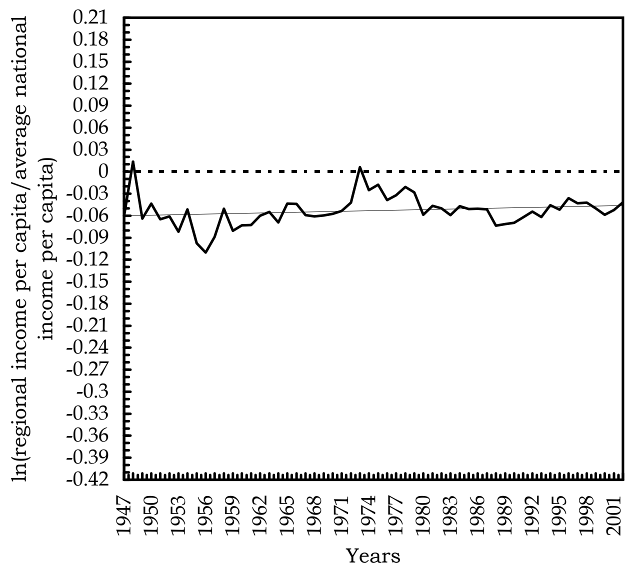


Figure 25. Southeast

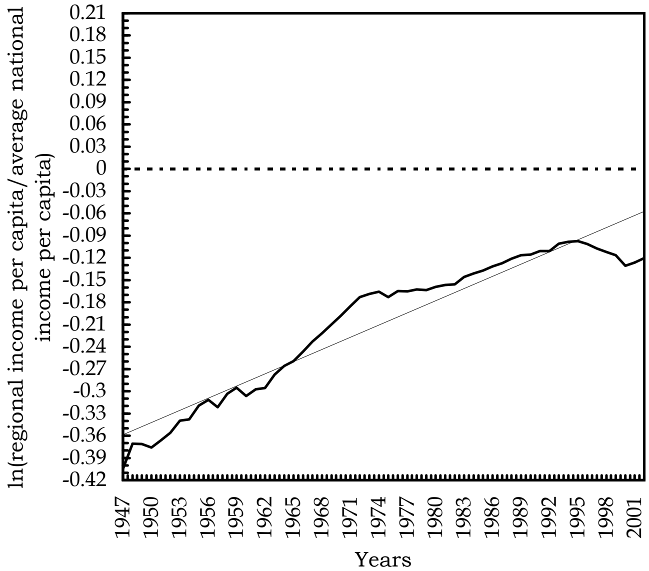


Figure 26. Southwest

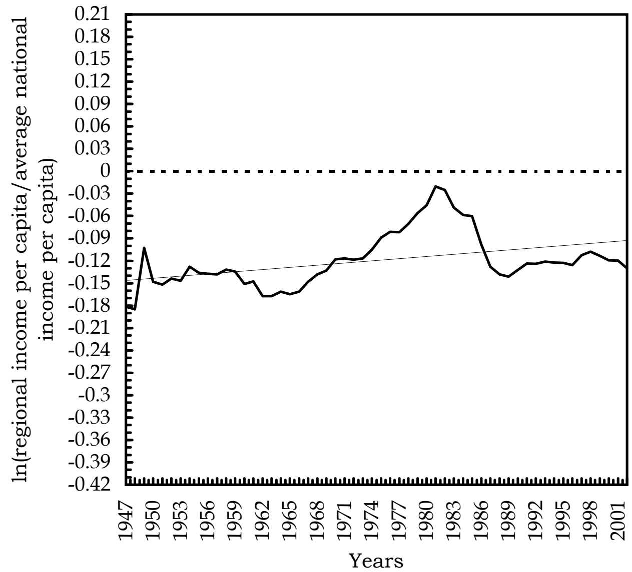


Figure 27. Rocky Mountains

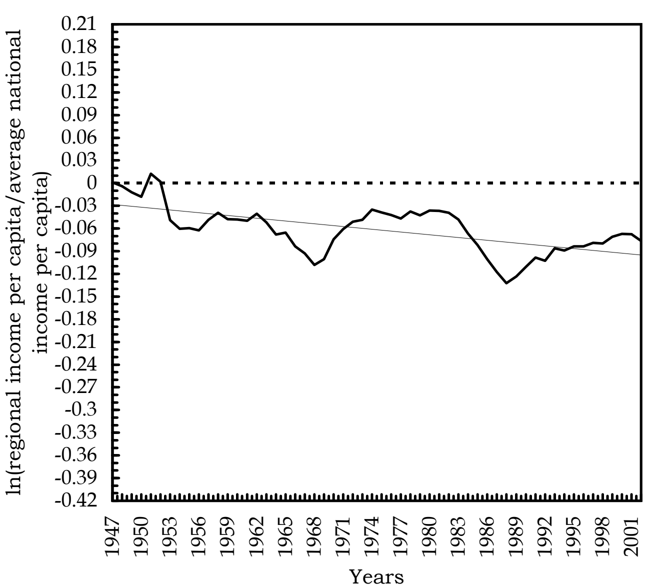


Figure 28. Farwest

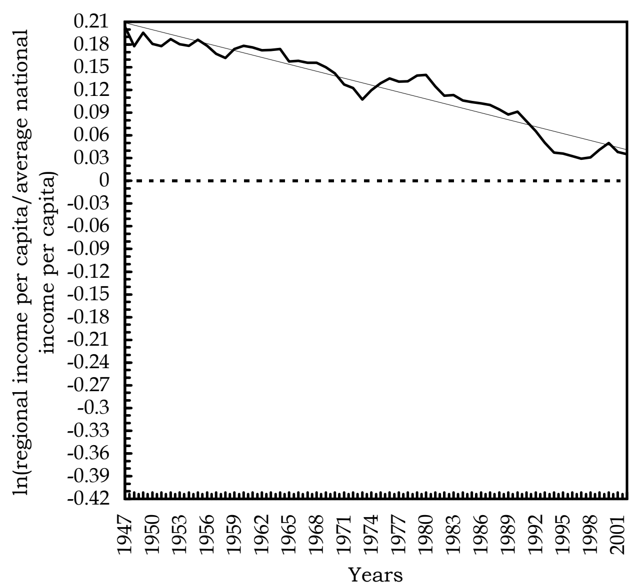


Figure 29. All U.S. Regions